1982

Normal Bayesian two-armed bandits

Steven K. Fahrenholtz

Iowa State University

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

Part of the Statistics and Probability Commons

Recommended Citation
Fahrenholtz, Steven K., "Normal Bayesian two-armed bandits " (1982). Retrospective Theses and Dissertations. 8343.
https://lib.dr.iastate.edu/rtd/8343

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

This reproduction was made from a copy of a document sent to us for microfilming. While the most advanced technology has been used to photograph and reproduce this document, the quality of the reproduction is heavily dependent upon the quality of the material submitted.

The following explanation of techniques is provided to help clarify markings or notations which may appear on this reproduction.

1. The sign or "target" for pages apparently lacking from the document photographed is "Missing Page(s)". If it was possible to obtain the missing page(s) or section, they are spliced into the film along with adjacent pages. This may have necessitated cutting through an image and duplicating adjacent pages to assure complete continuity.

2. When an image on the film is obliterated with a round black marl, it is an indication of either blurred copy because of movement during exposure, duplicate copy, or copyrighted materials that should not have been filmed. For blurred pages, a good image of the page can be found in the adjacent frame. If copyrighted materials were deleted, a target note will appear listing the pages in the adjacent frame.

3. When a map, drawing or chart, etc., is part of the material being photographed, a definite method of "sectioning" the material has been followed. It is customary to begin filming at the upper left hand corner of a large sheet and to continue from left to right in equal sections with small overlaps. If necessary, sectioning is continued again—beginning below the first row and continuing on until complete.

4. For illustrations that cannot be satisfactorily reproduced by xerographic means, photographic prints can be purchased at additional cost and inserted into your xerographic copy. These prints are available upon request from the Dissertations Customer Services Department.

5. Some pages in any document may have indistinct print. In all cases the best available copy has been filmed.
Fahrenholtz, Steven K.

NORMAL BAYESIAN TWO-ARMED BANDITS

Iowa State University Ph.D. 1982

University Microfilms International 300 N. Zeeb Road, Ann Arbor, MI 48106
Normal Bayesian two-armed bandits

by

Steven K. Fahrenholtz

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

Major: Statistics

Approved:
Signature was redacted for privacy.

In Charge of Major Work
Signature was redacted for privacy.

For the Major Department
Signature was redacted for privacy.

For the Graduate College

Iowa State University
Ames, Iowa

1982
TABLE OF CONTENTS

1. INTRODUCTION AND LITERATURE REVIEW ........................................ 1

2. INDEPENDENT NORMAL PRIOR ......................................................... 5
   2.1. Background ................................................................. 5
   2.2. Bayesian Distributions and Parametrization Under Normality ........... 6
   2.3. Bayesian Risk Equations .................................................... 8
   2.4. The Source-differential Function $\Delta^n(.)$ ......................... 18
   2.5. The Case $n=2$ ................................................................ 27
   2.6. Additional Properties of $\Delta^n(.)$ .................................... 35
   2.7. Bound on $\Delta^n(.)$ ...................................................... 49
   2.8. Approximation of $\Delta^n(.)$ ............................................ 59
   2.9. Numerical Computations .................................................... 87

3. SINGULAR NORMAL PRIOR .......................................................... 90
   3.1. Background .................................................................. 90
   3.2. Bayesian Distributions for the Singular Prior Case ..................... 90
   3.3. The Singular Prior Source Differential Function ....................... 94
   3.4. Equal Source Information ................................................... 100

4. BIBLIOGRAPHY ......................................................................... 103

5. ACKNOWLEDGMENTS ............................................................... 106
1. INTRODUCTION AND LITERATURE REVIEW

Two-armed bandit problems have their foundations in the realm of gambling and therefore can be best described in those terms. Consider a situation where

i) there are two slot machines labeled X and Y;

ii) \( F_X(\cdot \mid \theta) \) and \( F_Y(\cdot \mid \theta) \) are the distribution functions of the payoffs of machines X and Y, respectively, and

iii) \( \theta \) is unknown.

A strategy for a problem of this sort is a rule for deciding which machine to play next, based on the outcomes seen so far. The objective is to find a strategy that maximizes the discounted or undiscounted total payoffs.

Bandit problems have applications outside of the gambling context. The most evident application is to the sequential assignment of patients to treatments. The major difficulty in applying the bandit structure to such practical situations is that either the discount factor or the total number of patients to be treated must be known.

In general, bandit problems have two conflicting driving forces. The first is the need for obtaining information on the unknown parameter \( \theta \), while the second is the need for obtaining the largest possible immediate reward. This can be best illustrated by considering the following situation. Suppose that machine X appears to be the superior machine, but that significantly fewer observations have been taken on machine Y. The conflict arises in deciding when to sample the apparently
inferior machine to make sure that it is indeed inferior.

Generally, work in this area can be classified according to whether it is Bayesian or not, and also according to whether the cost function used places value on the correct identification of the state of nature (sequential design), or directly on the sample outcomes (bandit problems).

The bandit problem is first proposed in Robbins (1952) and one of the earliest Bayesian bandits is that considered by Feldman (1962). Feldman maximizes the expected winnings when faced with playing a symmetric Bernoulli two-armed bandit. The bandit is symmetric in the sense that a prior is placed on the points \((p_1, p_2), (p_2', p'_1)\) where \(p_1, p_2, p_1', p_2' \in (0,1)\). Kelly (1974) extends this work to arbitrary two point priors. Rodman (1978) considers the \(n\)-armed bandit where only one unknown arm is inferior. Berry (1978) uses Feldman's solution as the basis for a more general treatment allocation problem.

The Bernoulli two-armed bandit with arbitrary priors is discussed in Berry (1972). Berry proposes several conjectures which are proven in Gittins (1975) and Joshi (1975).

The sequential testing of the mean drift of a Wiener process is discussed in Breakwell and Chernoff (1964), Chernoff (1965a), and Chernoff (1965b). In Chernoff (1968), the Wiener Bayesian bandit is discussed. The main result of this paper shows how a two-armed Wiener bandit is similar to a one-armed bandit when a large amount of information has been gathered on one source. The two-armed normal discounted
bandit with an infinite stream of observations is mentioned in Gittins (1979). The basic result of this paper is a forward induction equation on the information.

A related non-Bayesian sequential design problem is discussed in Robbins and Siegmund (1974). This paper sequentially tests the hypothesis that one normal mean is larger than another while trying to reduce the number of observations on the inferior population. A non-Bayesian sequential probability ratio test is suggested. Their work is extended to cover the situation where the variances are unknown by Hayre and Gittins (1981).

This thesis is concerned with undiscounted two-armed Bayesian bandits where the number of trials is a known finite number. The variance of each source is assumed to be known and the objective is to find a strategy which minimizes the expected sum of the observations. The objective here differs slightly from the usual one in that we are trying to minimize the expected sum. This is strictly a matter of whether one wishes to work with a minimization or a maximization problem.

Chapter 2 is devoted to the situation where two independent normal priors are placed on the mean vector $(\theta_1, \theta_2)$ and the following structure is assumed:

i) two sources labeled X and Y;

ii) the payoffs for source X and Y are distributed as $N(\theta_1, \sigma^2_1)$ and $N(\theta_2, \sigma^2_2)$ random variables, respectively, where $\sigma^2_1$ and $\sigma^2_2$ are known;

iii) $(\theta_1, \theta_2)$ is unknown;
iv) the number of trials, $n$, is known; and 

v) $\theta_1$ and $\theta_2$ are independently distributed as $N(\mu_1, \tau_1^2)$ and $N(\mu_2, \tau_2^2)$ random variables, respectively, where $\mu_1$, $\tau_1^2$, $\mu_2$, and $\tau_2^2$ are known.

A description of the results of Chapter 2 can be found in Chapter 2, section 2.1.

Chapter 3 discusses the situation where the prior distribution of the mean vector $(\theta_1, \theta_2)$ is assumed to be normally distributed along a line. This situation will be referred to as the singular prior case. The singularity of the prior enables simple optimal strategies to be developed for certain special cases. The problem of Chapter 3 is a normal analog of Feldman (1962), while the problem of Chapter 2 is a partial analog of Berry (1972). We note that it might not be productive to extend the finite horizon total reward analysis of this dissertation to an infinite horizon LAEC (limiting average expected cost) analysis since this criterion seems to be too undemanding for the class of problems considered here.
2. INDEPENDENT NORMAL PRIOR

2.1. Background

In this chapter, we will investigate the problem of how to sequentially decide which of two independent normal sources to sample when the objective is to minimize the expected sum of the observations. Our attention will be focused on the case where we know in advance the variances of both sources and the total number \( n \) of observations that will be taken.

The second section reviews the normal posterior and marginal distributions when a normal prior is assumed. The next section precisely defines the loss function representing the expected sum of the observations, characterizes the optimal strategy, and develops recursive equations for the Bayes risk of each source. Section 2.4 investigates the source differential function \( \Delta^N(\cdot) \) and its reparametrizations. One of these reparametrizations shows that the assumption of known variances is less severe than it appears.

Section 2.5 motivates the theorems in the next three sections by studying the case where \( n=2 \). Theorems concerning the limiting behavior of \( \Delta^N(\cdot) \) and a bound for \( \Delta^N(\cdot) \) are provided in sections 2.6 and 2.7, respectively. Four linear approximations of \( \Delta^N(\cdot) \) are developed in section 2.8. The final section provides numerical computations of \( \Delta^N(\cdot) \), the bound for \( \Delta^N(\cdot) \), and the three approximations.
2.2. Bayesian Distributions and Parametrization
Under Normality

Consider the two information sources X and Y for the two-armed bandit problem of Chapter 1. Let x denote an observation from source X and y an observation from source Y.

Assume that

i) \( x \sim \mathcal{N}(\theta_1, \sigma_1^2); \sigma_1^2 > 0 \) \hspace{1cm} (2.2.1)

ii) \( y \sim \mathcal{N}(\theta_2, \sigma_2^2); \sigma_2^2 > 0 \)

iii) \( \sigma_1^2 \) and \( \sigma_2^2 \) are known

iv) \( x \) and \( y \) are independent.

The Bayesian approach calls for prior distributions on the unknown parameter pair \((\theta_1, \theta_2)\). A reasonable prior distribution is the independent normal. The main asset of this prior distribution is that for any combination of observations the posterior distribution of the means is again independent normal. This closure property allows recursive equations to be developed that characterize the optimal strategy.

Thus, the prior distributions on \((\theta_1, \theta_2)\) will satisfy

i) \( \theta_1 \sim \mathcal{N}(\mu_1, \tau_1^2); \tau_1^2 > 0 \) \hspace{1cm} (2.2.2)

ii) \( \theta_2 \sim \mathcal{N}(\mu_2, \tau_2^2); \tau_2^2 > 0 \)

iii) \( \theta_1 \) and \( \theta_2 \) independent,
and we are led to the following posterior and marginal distributions of both sources after an observation has been taken.

\begin{align*}
\text{i)} & \quad \theta_1 | x \sim N\left(\frac{\sigma_1^2}{\sigma_1^2 + \tau_1^2} \mu_1 + \frac{\tau_1^2}{\sigma_1^2 + \tau_1^2} x, \frac{\sigma_1^2 \tau_1^2}{\sigma_1^2 + \tau_1^2}\right) \\
\text{ii)} & \quad x \sim N\left(\mu_1, \sigma_1^2 + \tau_1^2\right) \\
\text{iii)} & \quad \theta_2 | y \sim N\left(\frac{\sigma_2^2}{\sigma_2^2 + \tau_2^2} \mu_2 + \frac{\tau_2^2}{\sigma_2^2 + \tau_2^2} y, \frac{\sigma_2^2 \tau_2^2}{\sigma_2^2 + \tau_2^2}\right) \\
\text{iv)} & \quad y \sim N\left(\mu_2, \sigma_2^2 + \tau_2^2\right) \\
\text{v)} & \quad \text{independence preserved in both posterior and marginal distributions.}
\end{align*}

Since information is additive, the natural parametrization turns out to be in terms of the information instead of the variances. In this parametrization, the relations 2.2.3 become

\begin{align*}
\text{i)} & \quad \theta_1 | x \sim N\left(\frac{I_1}{I_1 + J_1} \mu_1 + \frac{J_1}{I_1 + J_1} x, \frac{1}{I_1 + J_1}\right) \\
\text{ii)} & \quad x \sim N\left(\mu_1, \frac{1}{I_1} + \frac{1}{J_1}\right) \\
\text{iii)} & \quad \theta_2 | y \sim N\left(\frac{I_2}{I_2 + J_2} \mu_2 + \frac{J_2}{I_2 + J_2} y, \frac{1}{I_2 + J_2}\right) \\
\text{iv)} & \quad y \sim N\left(\mu_2, \frac{1}{I_2} + \frac{1}{J_2}\right) \\
\text{v)} & \quad \text{independence preserved in both posterior and marginal distributions.}
\end{align*}

This last parametrization will be used throughout the remainder
of the chapter. Relations 2.2.4 will be explicitly used in Theorem 2.3.2 which characterizes an optimal strategy and recursion Equations 2.3.7 and 2.3.8.

2.3. Bayesian Risk Equations

Up until now we have used the terms optimal strategy and expected sum in a rather informal manner. In this section, we will formalize these concepts and lay further groundwork.

The problem is to find a sequential strategy for taking observations so as to minimize the expected sum of the observations. Two concepts that need defining for the two-armed bandit problem are action space and strategy.

Definition 2.3.1: The action space, $A_n$, is given by

$$A_n = \{a_n : a_n = (a_1, a_2, \ldots, a_n)\}$$

where

$$a_i = \begin{cases} 
1 & \text{if source X is sampled at } i^{th} \text{ stage} \\
0 & \text{if source Y is sampled at } i^{th} \text{ stage}
\end{cases}$$

Definition 2.3.2: A strategy $s_n$ is a sequence of functions $s_n = (s_1, \ldots, s_n)$ where each $s_k$ maps the $k-1$ previous outcomes into $\{0,1\}$, yielding $a_k$.

Thus, a strategy is simply a rule that specifies which source to sample next, based on the outcomes observed so far. Since no previous
outcomes are available at the initial stage a strategy fixes the first source sampled.

A loss function representing the expected sum of the observations conditional on the composition \( \sum_{i=1}^{n} a_i, n - \sum_{i=1}^{n} a_i \) of the sample is

\[
L(\theta, a) = \theta_1 \sum_{i=1}^{n} a_i + \theta_2 (n - \sum_{i=1}^{n} a_i)
\]

where \( \theta = (\theta_1, \theta_2) \).

(2.3.1)

The risk \( R(\theta, s) \) and the Bayes risk \( R(\xi, s) \), for a strategy \( s \), need now be defined.

**Definition 2.3.3:**

\[
R(\theta, s) = \mathbb{E}(L(\theta, s))
\]

\[
= \theta_1 \mathbb{E}(\sum_{i=1}^{n} s_i | \theta) + \theta_2 \mathbb{E}(n - \sum_{i=1}^{n} s_i | \theta)
\]

Notice that for the problem of interest here, the expectation is with respect to the joint distribution of two independent normals with means and variances \( \theta_1, \frac{1}{J_1} \) and \( \theta_2, \frac{1}{J_2} \), respectively.

**Definition 2.3.4:**

\[
R(\xi, s) = \int R(\theta, s) d\xi(\theta)
\]

where \( \xi(\theta) \) is the prior distribution of \( \theta \).

As pointed out in Chapter 1, the prior here is assumed to be of the form

\[
\xi(\theta) = P(\theta_1)P(\theta_2)
\]

where \( P(\theta_1) \) and \( P(\theta_2) \) are the distribution functions of \( N(\mu_1, \frac{1}{I_1}) \)

and \( N(\mu_2, \frac{1}{I_2}) \) random variables, respectively.
The criterion that will be used to select the optimal strategy will be minimum Bayes risk.

**Definition 2.3.5:** An optimal strategy is any strategy, $s^*_n$, such that

$$R(\xi, s^*_n) = \inf_{s \in S^n} R(\xi, s_n)$$

where $S^n$ is the set of all strategies taking $n$ observations.

The dependence of the optimal strategy on the prior parameters $\mu_1, \mu_2, I_1, I_2$ is recognized by introducing

$$V_n(\mu_1, \mu_2, I_1, I_2) = \inf_{s \in S^n} R(\xi, s_n) \quad (2.3.2)$$

Notice that we have suppressed the dependence on $J_1$ and $J_2$, since these quantities are fixed for all values of $n$.

Two analogous quantities that will be useful in the characterization of the optimal strategy are

$$V^n_x(\mu_1, \mu_2, I_1, I_2) = \inf_{s \in S^n_x} R(\xi, s_n) \quad (2.3.3)$$

where $S^n_x = \{s_n: s_1 = 1\}$

and

$$V^n_y(\mu_1, \mu_2, I_1, I_2) = \inf_{s \in S^n_y} R(\xi, s_n) \quad (2.3.4)$$

where $S^n_y = \{s_n: s_1 = 0\}$.

Notice that $S^n_x$ is simply the set of all strategies that take an $X$ at
the first stages and a total of \( n \) observations and that \( S^n_y \) has a similar interpretation.

Since \( V^n_X(\cdot) \) and \( V^n_Y(\cdot) \) are defined to be infimums over two sets that form a partition of \( s^n \) we have the following lemma.

**Lemma 2.3.1**: \( V^n_X(\mu_1, \mu_2, I_1, I_2) = \min(V^n_X(\mu_1, \mu_2, I_1, I_2), V^n_Y(\mu_1, \mu_2, I_1, I_2)) \)

This lemma points in the direction of the optimal solution. It suggests that if the optimal Bayes risk for sampling \( X \) first is less than that for \( Y \) first it would make sense to sample \( X \); in other words one ought to sample the source having the smaller risk. The following theorem formalizes this concept.

**Theorem 2.3.2**: An optimal strategy is given by

1) sample source \( X \) if \( V^n_X(\mu_1^*, \mu_2^*, I_1^*, I_2^*) < V^n_Y(\mu_1^*, \mu_2^*, I_1^*, I_2^*) \)

2) sample source \( Y \) if \( V^n_Y(\mu_1^*, \mu_2^*, I_1^*, I_2^*) > V^n_Y(\mu_1^*, \mu_2^*, I_1^*, I_2^*) \)

3) sample either source if \( V^n_X(\mu_1^*, \mu_2^*, I_1^*, I_2^*) = V^n_Y(\mu_1^*, \mu_2^*, I_1^*, I_2^*) \)

where

\( n = \) number of observations still to be observed

\( \mu_1^* = \) current posterior mean of source \( X \)

\[ \mu_1^* = \frac{I_1}{I_1 + n_1 J_1} \mu_1 + \frac{n_1 J_1}{I_1 + n_1 J_1} x \]

\( n_1 = \) number of observations taken on source \( X \)
\( \bar{x} \) = mean of observations taken on source X

\( I_{1*} \) = current posterior information on source X

\[ I_{1*} = I_1 + n_1 J_1 \]

\( \mu_{2*} \) = current posterior mean of source Y

\[ \mu_{2*} = \frac{I_2}{I_2 + n_2 J_2} \mu_2 + \frac{n_2 J_2}{I_2 + n_2 J_2} \bar{y} \]

\( n_2 \) = number of observations taken on source Y

\( \bar{y} \) = mean of observations taken on source Y

\( I_{2*} \) = current posterior information on source Y

\[ I_{2*} = I_2 + n_2 J_2 \]

Recall that \( V^n_X(\cdot) \) and \( V^n_Y(\cdot) \) are simply the minimum Bayes risk when we are restricted to first sampling X and Y, respectively, and have to take a total of \( n \) observations. The optimal strategy of Theorem 2.3.2 is therefore seen to call for comparing \( V^n_X(\cdot) \) and \( V^n_Y(\cdot) \) as functions of the posterior parameters based on observations seen so far, with \( n \) equal to the number of observations still to be taken.

Both Feldman (1962) and Berry (1972) used the above idea, namely that the optimal strategy is given by sampling the source having the smaller current Bayes risk. Neither thought it necessary to provide any formal justification of the concept; such formal justification is provided in Blackwell and Girshick (1954) for certain related one source stopping problems. This thesis adopts the middle ground of providing a proof of Theorem 2.3.2 for only the case where \( n=2 \).
Proof of Theorem 2.3.2 for n=2:

Assume without loss of generality that

$$V_x^2(\mu_1, \mu_2, I_1, I_2) \leq V_y^2(\mu_1, \mu_2, I_1, I_2) \quad (2.3.5)$$

Case I: Let $s_2 = (s_1, s_2)$ be an arbitrary strategy such that $s_1=1$.

Therefore,

$$R(\xi, s_2) = \int \left[ E(\theta_1 + \theta_1 s_2(x) + \theta_2 (1-s_2(x)))dP(\theta_1)P(\theta_2) \right]$$

$$= \int (\theta_1 + \theta_2 + E((\theta_1 - \theta_2)s_2(x)))dP(\theta_1)P(\theta_2)$$

$$= \mu_1 + \mu_2 + \int E[(\theta_1 - \theta_2)s_2(x)]dP(\theta_1)P(\theta_2)$$

Now, if we interchange the order of integration using 2.2.4

we have that

$$R(\xi, s_2) = \mu_1 + \mu_2 + \int s_2(x) \left[ \int (\theta_1 - \theta_2) dP(\theta_1|x)P(\theta_2) \right] dP(x)$$

where $P(\theta_1|x)$ is the distribution function of a $N(\frac{I_1}{I_1+J_1} \mu_1 + \frac{J_1}{I_1+J_1} x,$ $\frac{1}{I_1+J_1}$) random variable, and $P(x)$ is the distribution function of a

$N(\mu_1', \frac{1}{I_1} + \frac{1}{J_1}$) random variable.

Therefore,

$$R(\xi, s_2) = \mu_1 + \mu_2 + \int (\mu_1(x) - \mu_2)s_2(x) dP(x) \quad (2.3.6)$$

where

$$\mu_1(x) = \text{posterior mean of source X given x}$$

$$= \frac{I_1}{I_1+J_1} \mu_1 + \frac{J_1}{I_1+J_1} x.$$
Let $s^*_2 = (s^*_1, s^*_2)$ denote the strategy given in Theorem 2.3.2. The strategy $s^*_2$ can be described as follows:

$$s^*_1 = 1$$

$$s^*_2 = \begin{cases} 
1 & \text{if } \mu_1(x) < \mu_2 \\
0 & \text{if } \mu_1(x) > \mu_2 
\end{cases}$$

Using 2.3.6 we have that

$$R(\xi, s^*_2) - R(\xi, s^*_2) = \int (\mu_1(x) - \mu_2)(s_2(x) - s^*_2(x))dP(x).$$

But

$$(\mu_1(x) - \mu_2)(s_2(x) - s^*_2(x)) \geq 0, \text{ for all } x,$$

since

$$\mu_1(x) - \mu_2 \geq 0 \Rightarrow s^*_2(x) = 0$$

$$\Rightarrow s_2(x) - s^*_2(x) \geq 0$$

and

$$\mu_1(x) - \mu_2 \leq 0 \Rightarrow s^*_2(x) = 1$$

$$\Rightarrow s_2(x) - s^*_2(x) \leq 0.$$

Therefore, $R(\xi, s^*_2) - R(\xi, s^*_2) \geq 0$ which implies that the Bayes risk for any strategy starting with source $X$ is larger than that of $s^*_2$.

Case II: Let $s_2 = (s_1, s_2)$ be an arbitrary strategy such that $s_1 = 0$. An argument similar to that given for Case I shows that

$$R(\xi, s_2) - R(\xi, s^*_2) \geq 0$$
where

\[ s^{**} = 0 \]

\[ s^{**}(y) = \begin{cases} 
1 & \text{if } \mu_1 < \mu_2(y) \\
0 & \text{if } \mu_1 > \mu_2(y)
\end{cases} \]

Therefore, all the strategies that start with source Y are dominated by \( s^{**} \). But,

\[
R(\xi, s^{**}) \geq V^2_y(\mu_1, \mu_2, I_1, I_2) \\
\geq V_x^2(\mu_1, \mu_2, I_1, I_2) \\
= R(\xi, s^*_2).
\]

The first and second inequalities follow by the definition of \( V^*_{y}(\cdot) \) and Equation 2.3.5, respectively. The last equality is a direct consequence of the analysis for Case I.

Therefore, we have shown that for \( n=2 \) the strategy \( s^*_2 \) given in Theorem 2.3.2 is optimal. Q.E.D.

Although the above theorem characterizes the optimal strategy, there are still large difficulties, since we need to compute \( V^*_{x}(\cdot) \) and \( V^*_{y}(\cdot) \) for all possible priors and \( n \). The task is made possible by recursion relations 2.3.7 and 2.3.8 below.

The Bayes risk \( V^n_x(\mu_1, \mu_2, I_1, I_2) \) is the average value of sampling \( X \) on the first trial plus the expectation of the anticipated Bayes risk, \( V^{n-1}_x \left( \frac{I_1}{I_1 + J_1} \mu_1 + \frac{J_1}{I_1 + J_1} x, \mu_2, I_1+J_1, I_2 \right) \), over the possible values
of \( x \). Therefore, the Bayes risk, \( V_x^{n}(\mu_1, \mu_2, I_1, I_2) \), must satisfy:

\[
V_x^{n}(\mu_1, \mu_2, I_1, I_2) = \mu_1 + \int \frac{1}{n-1} \left( \frac{I_1}{I_1+J_1} \right) \mu_1 + \frac{J_1}{I_1+J_1} x, \mu_2, I_1+J_1, I_2 \, dP(x)
\]

(2.3.7)

where \( P(x) \) is the distribution function of a \( N(\mu_1, \frac{1}{I_1} + \frac{1}{J_1}) \) random variable.

Similarly, the Bayes risk, \( V_y^{n}(\mu_1, \mu_2, I_1, I_2) \), must satisfy:

\[
V_y^{n}(\mu_1, \mu_2, I_1, I_2) = \mu_2 + \int \frac{1}{n-1} \left( \frac{I_2}{I_2+J_2} \right) \mu_2 + \frac{J_2}{I_2+J_2} y, \mu_1, I_1+J_2, I_2 \, dP(y)
\]

(2.3.8)

where \( P(y) \) is the marginal distribution function of a \( N(\mu_2, \frac{1}{I_2} + \frac{1}{J_2}) \) random variable.

In addition, one has

\[
V_x^{1}(\mu_1, \mu_2, I_1, I_2) = \mu_1
\]

and

\[
V_y^{1}(\mu_1, \mu_2, I_1, I_2) = \mu_2.
\]

Equation 2.3.7 and 2.3.8 are analogous to the recursions in Chernoff (1968), Feldman (1962), and Berry (1972).

Now, we will use Equations 2.3.7 and 2.3.8 to prove a certain symmetry relationship.

**Lemma 2.3.3:** \( V_y^{n}(\mu_1, \mu_2, I_1, I_2) = V_x^{n}(\mu_2, \mu_1, I_2, I_1) \), provided \( J_1 = J_2 = J \).
Proof: Let $n=1$, then

$$V^n_y(u_1, u_2, I_1, I_2) = u_2$$

$$= V^1_x(u_2, u_1, I_2, I_1).$$

Suppose the lemma holds for $n$; we need to show that it holds for $n+1$.

The induction hypothesis implies that

$$V^n(u_1, u_2, I_1, I_2) = V^n(u_2, u_1, I_2, I_1)$$

since

$$V^n(u_1, u_2, I_1, I_2) = \min(V^n_x(u_1, u_2, I_1, I_2), V^n_y(u_1, u_2, I_1, I_2))$$

$$= \min(V^n_y(u_2, u_1, I_2, I_1), V^n_x(u_2, u_1, I_2, I_1))$$

$$= V^n(u_2, u_1, I_2, I_1).$$

Now, using Equations 2.3.7 and 2.3.8 we have that

$$V^{n+1}_y(u_1, u_2, I_1, I_2) = u_2 + \int V^n(u_1, \frac{I_2}{I_2+J} u_2 + \frac{J}{I_2+J} y, I_1, I_2+J) dP(y)$$

$$= u_2 + \int V^n(u_1, \frac{I_2}{I_2+J} u_2 + \frac{J}{I_2+J} y, u_1, I_2+J, I_1) dP(y)$$

$$= V^{n+1}_x(u_2, u_1, I_2, I_1).$$

Corollary 2.3.4: $V^n(u_1, u_2, I_1, I_2) = V^n(u_2, u_1, I_1, I_2)$,

provided $J_1 = J_2 = J$.  

2.4. The Source-differential Function \( \Delta^n(\cdot) \)

Theorem 2.3.2 shows that the optimal strategy depends only on the sign of

\[
\Delta^n(\mu_1, \mu_2, I_1, I_2) = V^n_y(\mu_1, \mu_2, I_1, I_2) - V^n_x(\mu_1, \mu_2, I_1, I_2); \tag{2.4.1}
\]

so its essentials may be restated as follows:

i) sample source X if \( \Delta^n(\mu_1, \mu_2, I_1, I_2) > 0 \)

ii) sample source Y if \( \Delta^n(\mu_1, \mu_2, I_1, I_2) < 0 \)

iii) sample either source if \( \Delta^n(\mu_1, \mu_2, I_1, I_2) = 0 \).

This suggests that \( \Delta^n(\mu_1, \mu_2, I_1, I_2) \) is a quantity that should receive further attention.

A functional relationship similar to those of Equations 2.3.7 and 2.3.8, and motivated by Lemma 2.1 of Feldman (1962), can be proven as follows.

**Theorem 2.4.1:**

\[
\Delta^n(\mu_1, \mu_2, I_1, I_2) = \int \Delta^{n-1} \left( \frac{I_1}{I_1 + J_1} \mu_1 + \frac{J_1}{I_1 + J_1} x, \mu_2, I_1 + J_1, I_2 \right)^+ \, dP(x)
\]

\[
+ \int \Delta^{n-1} \left( \mu_1', \frac{I_2}{I_2 + J_2} \mu_2 + \frac{J_2}{I_2 + J_2} y, I_1, I_2 + J_2 \right)^- \, dP(y)
\]

where, with \( 1(E) \) the indicator of the event \( E \),
\[ \Delta^{n}(\mu_{1}, \mu_{2}, I_{1}, I_{2})^+ = \Delta^{n}(\mu_{1}, \mu_{2}, I_{1}, I_{2})_1 (\Delta^{n}(\mu_{1}, \mu_{2}, I_{1}, I_{2}) > 0) \]

\[ \Delta^{n}(\mu_{1}, \mu_{2}, I_{1}, I_{2})^- = \Delta^{n}(\mu_{1}, \mu_{2}, I_{1}, I_{2})_1 (\Delta^{n}(\mu_{1}, \mu_{2}, I_{1}, I_{2}) < 0) \]

\[ P(x) = \text{distribution function of a } \mathcal{N}(\mu_{1}, \frac{1}{I_{1}} + \frac{1}{J_{1}}) \]

\[ P(y) = \text{distribution function of a } \mathcal{N}(\mu_{2}, \frac{1}{I_{2}} + \frac{1}{J_{2}}) \]

**Proof:** Define

\[ V_{xy}^{n}(\mu_{1}, \mu_{2}, I_{1}, I_{2}) = \min_{s \in S_{xy}^{n}} R(\xi, s_{n}) \]

where

\[ S_{xy}^{n} = \{ s_{n} : s_{1} = 1 \text{ and } s_{2} = 0 \} \]

Define as well the analogous quantity

\[ V_{yx}^{n}(\mu_{1}, \mu_{2}, I_{1}, I_{2}) = \min_{s \in S_{yx}^{n}} R(\xi, s_{n}) \]

where

\[ S_{yx}^{n} = \{ s_{n} : s_{1} = 0 \text{ and } s_{2} = 1 \} \]

Notice that

\[ V_{xy}^{n}(\mu_{1}, \mu_{2}, I_{1}, I_{2}) = V_{yx}^{n}(\mu_{1}, \mu_{2}, I_{1}, I_{2}) \]  \hspace{1cm} (2.4.2)

since the order of receiving information is irrelevant.

Also, we have that

\[ V_{xy}^{n}(\mu_{1}, \mu_{2}, I_{1}, I_{2}) = v_{1} + \int_{Y} v^{n-1} \left( \frac{I_{1}}{I_{1} + J_{1}} \mu_{1} + \frac{J_{1}}{I_{1} + J_{1}} x, \mu_{2}, I_{1} + J_{1}, I_{2} \right) dP(x) \]  \hspace{1cm} (2.4.3)
Therefore, using Equation 2.4.2, we have that

\[ \Delta^n(\mu_1, \mu_2, I_1, I_2) = \nabla_x^n(\mu_1, \mu_2, I_1, I_2) - \nabla_y^n(\mu_1, \mu_2, I_1, I_2) \]

Now, after substituting Equations 2.3.7, 2.3.8, 2.4.3, 2.4.4 into Equation 2.4.5 we have that

\[
\Delta^n(\mu_1, \mu_2, I_1, I_2) = \int (\nabla_x^n(\mu_1, \frac{I_2}{I_2 + J_2} \mu_2 + \frac{J_2}{I_2 + J_2} y, I_1, I_2 + J_2) - \\
\nabla_y^n(\mu_1, \frac{I_2}{I_2 + J_2} \mu_2 + \frac{J_2}{I_2 + J_2} y, I_1, I_2 + J_2)) dP(y) \\
+ \int (\nabla_y^n(\frac{I_1}{I_1 + J_1} \mu_1 + \frac{J_1}{I_1 + J_1} x, \mu_2, I_1 + J_1, I_2) - \\
\nabla_x^n(\frac{I_1}{I_1 + J_1} \mu_1 + \frac{J_1}{I_1 + J_1} x, \mu_2, I_1 + J_1, I_2)) dP(x).
\]

Lemma 2.3.1 implies that

\[
\Delta^n(\mu_1, \mu_2, I_1, I_2) = \int \Delta^n(\mu_1, \frac{I_2}{I_2 + J_2} \mu_2 + \frac{J_2}{I_2 + J_2} y, I_1, I_2 + J_2) dP(y) \\
+ \int \Delta^n(\frac{I_1}{I_1 + J_1} \mu_1 + \frac{J_1}{I_1 + J_1} x, I_1 + J_1, I_2) dP(x).
\]

Q.E.D.
A direct consequence of Lemma 2.3.3 will now be stated.

Lemma 2.4.2: \( \Delta^n(\mu_1, \mu_2, I_1, I_2) = -\Delta^n(\mu_2, \mu_1, I_1, I_2) \), provided \( J_1 = J_2 = J \).

Proof:

\[
\Delta^n(\mu_1, \mu_2, I_1, I_2) = v^n_y(\mu_1, \mu_2, I_1, I_2) - v^n_x(\mu_1, \mu_2, I_1, I_2) \\
= v^n_x(\mu_2, \mu_1, I_2, I_1) - v^n_y(\mu_2, \mu_1, I_2, I_1) \\
= -\Delta^n(\mu_2, \mu_1, I_2, I_1).
\]

\( \Delta^n(\cdot) \) is a cumbersome function, since it is a function of seven arguments; \( \mu_1, \mu_2, I_1, I_2, J_1, J_2 \) and \( n \). Reparametrization reduces the number of arguments that need to be considered to four.

The first reparametrization results from the fact that \( \Delta^n(\cdot) \) is a function of the differences of the means. Therefore, we need not use \( \mu_1 \) and \( \mu_2 \) as arguments; only their difference \( \delta \).

Theorem 2.4.3: If \( \mu_2' - \mu_1' = \mu''_2 - \mu''_1 \), then

\[
\Delta^n(\mu_1', \mu_2', I_1, I_2) = \Delta^n(\mu''_1, \mu''_2, I_1, I_2).
\]

Proof:

Let \( n=1 \); then

\[
\Delta^1(\mu_1', \mu_2', I_1, I_2) = \mu_2' - \mu_1' \\
= \mu''_2 - \mu''_1 \\
= \Delta^1(\mu''_1, \mu''_2, I_1, I_2).
\]
Suppose the theorem holds for \( n \); we need to show that it holds for \( n+1 \).

Theorem 2.4.1 provides the following equality

\[
\Delta^{n+1}(\mu_1', \mu_2', I_1, I_2) = \int \Delta^n \left( \frac{I_1}{I_1+J_1} \mu_1' + \frac{J_1}{I_1+J_1} x, \mu_2', I_1+J_1, I_2 \right)^+ dP(x) \\
+ \int \Delta^n \left( \frac{I_2}{I_2+J_2} \mu_2' + \frac{J_2}{I_2+J_2} y, I_1, I_2+J_2 \right)^- dP(y)
\]

where \( P(x) \) and \( P(y) \) are the distribution functions of \( N(\mu_1', \frac{1}{I_1} + \frac{1}{J_1}) \) and \( N(\mu_2', \frac{1}{I_2} + \frac{1}{J_2}) \) random variables.

Now, we must standardize the two marginal distributions. Let

\[
x = \mu_1' + \left( \frac{I_1+J_1}{I_1 J_1} \right)^{1/2} z_1
\]

and

\[
y = \mu_2' + \left( \frac{I_2+J_2}{I_2 J_2} \right)^{1/2} z_2
\]

where \( z_1 \) and \( z_2 \) are standard normal random variables.

Therefore,

\[
\Delta^{n+1}(\mu_1', \mu_2', I_1, I_2) = \int \Delta^n \left( \frac{J_1}{I_1(I_1+J_1)} \mu_1' + \left( \frac{1}{I_1(I_1+J_1)} \right)^{1/2} z_1, \mu_2', I_1+J_1, I_2 \right)^+ d\Phi(z) \\
+ \int \Delta^n \left( \frac{J_2}{I_2(I_2+J_2)} \mu_2' + \left( \frac{1}{I_2(I_2+J_2)} \right)^{1/2} z_2, I_1, I_2+J_2 \right)^- d\Phi(z)
\]

(2.4.6)
The induction hypothesis implies that

\[
\Delta^{n+1}(\mu_1', \mu_2', I_1, I_2) = \int \Delta^n(\mu''_1 + (\frac{J_1}{I_1(I_1+J_1)})^{1/2}z_1, \mu''_2, I_1, I_1+J_1, I_2)^+d\phi(z)
+ \int \Delta^n(\mu''_1, \mu''_2 + (\frac{J_2}{I_2(I_2+J_2)})^{1/2}z_2, I_1, I_1+I_2)^-d\phi(z).
\]

Therefore,

\[
\Delta^{n+1}(\mu_1', \mu_2', I_1, I_2) = \int \Delta^n(\frac{I_1}{I_1+J_1} \mu''_1 + \frac{J_1}{I_1+J_1} x, \mu''_2, I_1, I_1+J_1, I_2)^+d\phi(x)
+ \int \Delta^n(\mu''_1, \frac{I_2}{I_2+J_2} \mu''_2 + \frac{J_2}{I_2+J_2} y, I_1, I_2+I_2)^-d\phi(y)
= \Delta^{n+1}(\mu''_1, \mu''_2, I_1, I_2).
\]

Q.E.D.

**Corollary 2.4.4:**

\[
\Delta^n(\mu_1, \mu_2, I_1, I_2) = \Delta^n(\mu_1-\mu_2, 0, I_1, I_2) = \Delta^n(0, \mu_2-\mu_1, I_1, I_2)
\]

The above corollary implies that \(\Delta^n(\cdot)\) is a function of the difference of means. Therefore, \(\Delta^n(\cdot)\) may be written as

\[
\Delta^n(\delta, I_1, I_2), \text{ with } \delta = \mu_2-\mu_1.
\] (2.4.7)

Notice that if we use this expression in conjunction with the standardization of 2.4.6, we can rewrite the result of Theorem 2.4.1 as

\[
\Delta^{n+1}(\delta, I_1, I_2) = \int \Delta^n(\delta-z(\frac{J_1}{I_1(I_1+J_1)})^{1/2}, I_1+J_1, I_2)^+d\phi(z)
= \int \Delta^n(\delta+z(\frac{J_2}{I_2(I_2+J_2)})^{1/2}, I_1, I_2+J_2)^-d\phi(z)
\] (2.4.8)
Up until now we have suppressed the role of $J_1$ and $J_2$, the inverses of the known variances, since they are fixed quantities for all values of $n$. In the next theorem, further reparametrization is accomplished by showing that $\Delta^n(\cdot)$ depends $I_1, I_2, J_1$, and $J_2$ only through the ratios $\frac{I_1}{J_1}$ and $\frac{I_2}{J_2}$. This may be interpreted as showing that $I_i$ is naturally measured in units of $J_i$.

Since in the following theorem we will be dealing explicitly with the quantities $J_1$ and $J_2$, we will for the time being add them to the argument list of $\Delta^n(\cdot)$.

Theorem 2.4.5: If

$$I_1' = \frac{I_1''}{J_1'} \quad \text{and} \quad I_2' = \frac{I_2''}{J_2'},$$

(2.4.9)

then

$$\Delta^n(\delta, I_1', I_2', J_1', J_2') = \Delta^n(\delta, I_1'', I_2'', J_1'', J_2'').$$

(2.4.10)

Proof: Let $n=1$, then

$$\Delta^1(\delta, I_1', I_2', J_1', J_2') = \delta$$

$$= \Delta^1(\delta, I_1', I_2', J_1', J_2').$$

Let the conclusion of the theorem hold for $n$; we need to show that it holds for $n+1$.

Equation 2.4.8 implies that
\[ \Delta^{n+1}(\delta, I_1', I_2', J_1', J_2') = \int \Delta^n(\delta - z((\frac{J_1^l}{J_1^e}) + 1))^{-1/2}, I_1', J_1', I_2', J_2')^+d\phi(z) \]

\[ + \int \Delta^n(\delta + z((\frac{J_1^l}{J_1^e}) + 1))^{-1/2}, I_1', I_2', J_1', J_2')^+d\phi(z). \]

But, by Equation 2.4.9 we have that

\[ \frac{I_1^l}{J_1^e} \left( \frac{J_1^l}{J_1^e} + 1 \right) = \left( \frac{I_1^l}{J_1^e} \right)^+ \]

and

\[ \frac{I_2^l}{J_2^e} \left( \frac{J_2^l}{J_2^e} + 1 \right) = \left( \frac{I_2^l}{J_2^e} \right)^+ \]

so that

\[ \Delta^{n+1}(\delta, I_1', I_2', J_1', J_2') = \int \Delta^n(\delta - z((\frac{I_1^l}{J_1^e}) + 1))^{-1/2}, I_1', I_2', J_1', J_2')^+d\phi(z) \]

\[ + \int \Delta^n(\delta + z((\frac{I_2^l}{J_2^e}) + 1))^{-1/2}, I_1', I_2', J_1', J_2')^+d\phi(z). \]

(2.4.11)

Therefore, by the induction hypotheses, we have that

\[ \Delta^n(\delta - z((\frac{I_1^l}{J_1^e}) + 1))^{-1/2}, I_1', I_2', J_1', J_2')^+ \]

\[ = \Delta^n(\delta - z((\frac{I_1^l}{J_1^e}) + 1))^{-1/2}, I_1^l, I_2^l, J_1^l, J_2^l) \]

and

\[ \Delta^n(\delta + z((\frac{I_2^l}{J_2^e}) + 1))^{-1/2}, I_1', I_2', J_1', J_2')^+ \]

\[ = \Delta^n(\delta + z((\frac{I_2^l}{J_2^e}) + 1))^{-1/2}, I_1^l, I_2^l, J_1^l, J_2^l) \]

in view of Equation 2.4.9.

Hence, Equation 2.4.11 can be written as
\[ \Delta^{n+1}(\delta, I_1', I_2', J_1', J_2') = \int \Delta^n(\delta - z(x_1, x_2))^{-1/2} \, d\phi(z) \]

\[ + \int \Delta^n(\delta + z(x_1, x_2))^{-1/2} \, d\phi(z) \]

Q.E.D.

**Corollary 2.4.6:**

\[ \Delta^n(\delta, I_1, I_2, J_1, J_2) = \Delta^n(\delta, \frac{I_1}{J_1}, \frac{I_2}{J_2}, 1, 1) \]

In view of Theorem 2.4.5 we will henceforth write \( \Delta^n(\cdot) \) in the form \( \Delta^n(\delta, I_1, I_2) \) where \( I_1 \) and \( I_2 \) are understood to represent the ratios \( \frac{I_1}{J_1} \) and \( \frac{I_2}{J_2} \), respectively.

Notice that if we combine Equation 2.4.8 and Theorem 2.4.5 we have the fully reparametrized version of Theorem 2.4.1,

\[ \Delta^n(\delta, I_1', I_2') = \int \Delta^n(\delta - z(I_1 + 1), I_1, I_2) \, d\phi(z) \]

\[ + \int \Delta^n(\delta + z(I_2 + 1), I_1, I_2) \, d\phi(z). \quad (2.4.12) \]

Theorem 2.4.5 implies that we need only let the change in information on a source from one stage to the next be one if an observation is taken from that source and zero otherwise. Therefore, the information on a particular source at any stage is the number of times that source has been sampled plus \( n^*_i = \frac{I^*_i}{J_i} \), where \( I^*_i \) is the initial information on source \( i \). The quantity \( n^*_i \) represents how many observations worth of information we have on source \( i \) before we start sampling.
Now, we can restate Theorem 2.3.2 in the following form:

An optimal strategy is given by

i) sample source X if $\Delta^n(\delta, I_1, I_2) > 0$

ii) sample source Y if $\Delta^n(\delta, I_1, I_2) < 0$

iii) sample either source if $\Delta^n(\delta, I_1, I_2) = 0$

where

\[ n = \text{number of observations still to be observed} \]
\[ \delta = \text{current difference in posterior means} \]
\[ I_1 = n_1 + n^*_1 \]
\[ I_2 = n_2 + n^*_2 \]
\[ n_1 = \text{number of observations on X} \]
\[ n_2 = \text{number of observations on Y} \]

In some situations, we are not willing to assume any prior information on either source. Improper priors on both sources is one solution to this problem. If we let $I_1^* = I_2^* = 0$, and proceed to take one observation from each source, we will have proper priors for all the remaining stages. Also, we would have the following simplifications; $\delta = y-x$, $I_1 = n_x$, and $I_2 = n_y$.

2.5. The Case n=2

In this section, we focus our attention on the case where we are going to take two observations. The reason that the situation where n=2 is of special interest is that it will motivate several theorems found in the next section and be the basis for several
approximations of $\Delta^h(\cdot)$.

Using Equation 2.4.12, we have that

$$\Delta^2(\delta, I_1, I_2) = \int \Delta^1(\delta - z((I_1(I_1+1))^{-1/2}, I_1+1, I_2)^+d\Phi(z)$$
$$+ \int \Delta^1(\delta + z((I_2(I_2+1))^{-1/2}, I_1, I_2+1)^-d\Phi(z)$$
$$= \int ((\delta - z(I_1(I_1+1))^{-1/2})^+d\Phi(z)$$
$$+ \int ((\delta + z(I_2(I_2+1))^{-1/2})^-d\Phi(z)$$

Now,

$$\int_{-\infty}^t zd\Phi(z) = -\Phi(t)$$

implies that

$$\Delta^2(\delta, I_1, I_2) = \delta(\delta(I_1(I_1+1))^{1/2}) + \Phi(-\delta(I_2(I_2+1))^{1/2})$$
$$+ (I_1(I_1+1))^{-1/2}\Phi(\delta(I_1(I_1+1))^{1/2})$$
$$- (I_2(I_2+1))^{-1/2}\Phi(-\delta(I_2(I_2+1))^{1/2}). \quad (2.5.1)$$

Now, let us investigate $\Delta^2(\cdot)$ as a function of $\delta$ for fixed values of $I_1$ and $I_2$. To begin with, for fixed values of $I_1$ and $I_2$, compute

$$\frac{d}{d\delta} \Delta^2(\delta, I_1, I_2) = \delta(\delta(I_1(I_1+1))^{1/2}) + \Phi(-\delta(I_2(I_2+1))^{1/2})$$
$$+ \delta[(I_1(I_1+1))^{1/2}\Phi(\delta(I_1(I_1+1))^{1/2}) - (I_2(I_2+1))^{1/2}\Phi(-\delta(I_2(I_2+1))^{1/2})]$$
$$- (I_1(I_1+1))^{-1/2}\Phi(\delta(I_1(I_1+1))^{1/2}) (I_1(I_1+1))\delta$$
$$+ (I_2(I_2+1))^{-1/2}\Phi(-\delta(I_2(I_2+1))^{1/2}) (I_2(I_2+1))\delta$$
$$= \Phi(\delta(I_1(I_1+1))^{1/2}) + \Phi(-\delta(I_2(I_2+1))^{1/2}) \quad (2.5.2)$$
The following lemma will provide the motivation for the theorems found in sections 2.6 and 2.7.

**Lemma 2.5.1:**

i) $\Delta^{2}(\delta, I_1, I_2)$ is an increasing function in $\delta$, for all $I_1, I_2$

ii) $\lim_{\delta \to \infty} \frac{\Delta^{2}(\delta, I_1, I_2)}{\delta} = 1$, $\lim_{\delta \to \infty} \frac{\Delta^{2}(\delta, I_1, I_2)}{\delta} = 1$

iii) $\lim_{\min(I_1, I_2) \to \infty} \Delta^{2}(\delta, I_1, I_2) = \delta$

**Proof:**

i) $\frac{d}{d\delta} \Delta^{2}(\delta, I_1, I_2) > 0$, for all $I_1, I_2$.

ii) $\lim_{\delta \to \infty} \frac{\Delta^{2}(\delta, I_1, I_2)}{\delta} = \lim_{\delta \to \infty} [\Phi(\delta(I_1(I_1+1))^{1/2}) + \Phi(-\delta(I_2(I_2+1))^{1/2})] = 1$.

A similar argument is used for $\delta \to \infty$.

iii) $\lim_{\min(I_1, I_2) \to \infty} \Delta^{2}(\delta, I_1, I_2) = \lim_{\min(I_1, I_2) \to \infty} \delta(\Phi(\delta(I_1(I_1+1))^{1/2}) + \Phi(-\delta(I_2(I_2+1))^{1/2}))$

Now, if $\delta > 0$, then $\Phi(\delta(I_1(I_1+1))^{1/2}) \to 1$ and $\Phi(-\delta(I_2(I_2+1))^{1/2}) \to 0$,

if $\delta < 0$, then $\Phi(\delta(I_1(I_1+1))^{1/2}) \to 0$ and $\Phi(-\delta(I_2(I_2+1))^{1/2}) \to 1$,

if $\delta = 0$, then $\delta(\Phi(\delta(I_1(I_1+1))^{1/2}) + \Phi(-\delta(I_2(I_2+1))^{1/2})) = 0$. 


Hence,
\[
\lim_{\min(I_1, I_2) \to \infty} \Delta^2(\delta, I_1, I_2) = \delta, \text{ for all } \delta.
\]
Q.E.D.

The following theorem will show that \(\Delta^2(\cdot)\) is a decreasing function in \(I_1\) and an increasing function in \(I_2\).

**Theorem 2.5.2:**
\[\Delta^2(\delta, I_1, I_2)\] is a [decreasing] function in \([I_1^1]^{I_1}\) and an [increasing] function in \([I_2^2]^{I_2}\).

**Proof:**
\[
\Delta^2(\delta, I_1, I_2) = \int (\delta-z(I_1(I_1+1))^{-1/2})^++d\Phi(z)
+ \int (\delta+z(I_2(I_2+1))^{-1/2})^++d\Psi(z)
= A(\delta, I_1) + B(\delta, I_2).
\]

First, we will prove that \(\Delta^2(\delta, I_1, I_2)\) is a decreasing function in \(I_1\) which is equivalent to showing that \(A(\delta, I_1)\) is a decreasing function in \(I_1\) since \(B(\cdot)\) is free of \(I_1\).

Let \(I_1' < I_1''\); we need to show that \(A(\delta, I_1') - A(\delta, I_1'') > 0\).

Define
\[
k'(z) = \delta-z(I_1'(I_1'+1))^{-1/2}
\]
\[
k''(z) = \delta-z(I_1''(I_1''+1))^{-1/2}.
\]
Case I: \( \delta > 0 \)

Notice that

\[ a' < a'' \]

where

\[ a' = \delta (I_1''(I_1'' + 1))^{1/2} \]

and

\[ a'' = \delta (I_1''(I_1'' + 1))^{1/2}. \]  \hspace{1cm} (2.5.3)

Therefore, \( k'(z) \) and \( k''(z) \) may be pictured as in Figure 2.5.1.

Now,

\[
0 = \int (k'(z) - k''(z)) \, d\phi(z)
\]

\[
= \int_{-\infty}^{a'} (k'(z) - k''(z)) \, d\phi(z) + \int_{a'}^{a''} (k'(z) - k''(z)) \, d\phi(z)
\]

\[
+ \int_{a''}^{\infty} (k'(z) - k''(z)) \, d\phi(z)
\]

\[
< \int_{-\infty}^{a'} (k'(z) - k''(z)) \, d\phi(z) + \int_{a'}^{a''} (k'(z) - k''(z)) \, d\phi(z)
\]

\[
< \int_{-\infty}^{a'} k'(z) \, d\phi(z) - \int_{-\infty}^{a''} k''(z) \, d\phi(z)
\]

\[
= A(\delta, I_1') - A(\delta, I_1'')
\]

The first equality follows since \( E(z) = 0 \) while Equation 2.5.3 implies the next equality. The next two inequalities follow from Figure 2.5.1.
Figure 2.5.1. Plot of the lines $k'(z)$ and $k''(z)$ for the case $\delta > 0$
Case II: $\delta < 0$

Notice that

$$a'' < a'$$

where $a''$ and $a'$ are defined as in Equation 2.5.3. Therefore, $k'(z)$ and $k''(z)$ may be pictured as in Figure 2.5.2.

Figure 2.5.2 shows that

$$k'(z)1(k'(z) > 0) > k''(z)1(k''(z) > 0).$$

Therefore,

$$0 < \int (k'(z)^+ - k''(z)^+)d\phi(z) = A(\delta, I_1') - A(\delta, I_1'')$$

Hence, we have proven that $A^2(\delta, I_1, I_2)$ is a decreasing function in $I_1$.

It is now easy to prove that $A^2(\delta, I_1, I_2)$ is an increasing function in $I_2$, since by Lemma 2.4.2, we have that

$$\Delta^2(\delta, I_1, I_2) = -\Delta^2(-\delta, I_2, I_1).$$

Q.E.D.

The propositions of Lemma 2.5.1 can be extended to an arbitrary $n$, and will be so extended in section 2.6. Unfortunately, we do not have a proof of Theorem 2.5.2 for any $n$ greater than two. The main difficulty is showing that

$$\int A^n(\delta + z(I_2 + 1)^{-1/2}, I_1, I_2 + 1)^{-d\phi(z)}$$

is a nonincreasing function of $I_1$ when $n \geq 2$. Notice that this was not a problem when $n = 2$ since
Figure 2.5.2. Plot of the lines $k'(z)$ and $k''(z)$ for the case $\delta < 0$
\[ \int \Delta^1(\delta + z(I_2(I_2+1))^{-1/2}, I_1, I_2+1)^{-\delta} \, \Phi(z) \]
does not depend on \( I_1 \).

### 2.6. Additional Properties of \( \Delta^N(\cdot) \)

This section will be devoted to showing that \( \Delta^N(\cdot) \) is a uniformly continuous increasing function of \( \delta \), and to investigating the limiting behavior of \( \Delta^N(.) \) with respect to the informations \( I_1, I_2 \) and \( \delta \).

Most of the theorems proven in this section are direct extensions of the results of section 2.5. Theorem 2.6.2 establishes uniform continuity of \( \Delta^N(\cdot) \). Theorem 2.6.4 bounds \( \Delta^N(\cdot) \). This bound is then used in Theorems 2.6.5 and 2.6.6, respectively, extending parts (ii) and (iii) of Lemma 2.5.1 to the case of arbitrary \( n \). Finally, Theorem 2.6.6 leads to Theorem 2.6.7, extending part (i) of Lemma 2.5.1 to arbitrary \( n \).

The following lemma is needed in the proof of Theorem 2.6.2.

**Lemma 2.6.1:**

1. \[ |f(x)^+ - g(x)^+| \leq |f(x) - g(x)| \]
2. \[ |f(x)^- - g(x)^-| \leq |f(x) - g(x)| \]

where \( f(x)^+ = f(x) 1(f(x) > 0) \) and \( f(x)^- = f(x) 1(f(x) \leq 0) \).

**Proof of i:**

Case 1: \( f(x) > 0 \) and \( g(x) > 0 \) \( \Rightarrow \) \[ |f(x)^+ - g(x)^+| = |f(x) - g(x)| \]
Case 2: \( f(z) > 0 \) and \( g(x) \leq 0 \) \( \Rightarrow \) \( |f(x) + g(x)^+| = |f(x)| \leq |f(x) - g(x)| \)

Case 3: \( f(x) < 0 \) and \( g(x) > 0 \) \( \Rightarrow \) \( |f(x) + g(x)^+| = |g(x)| \leq |g(x) - f(x)| \)

Case 4: \( f(x) < 0 \) and \( g(x) < 0 \) \( \Rightarrow \) \( |f(x) + g(x)^+| = 0 \)

\[ \leq |f(x) - g(x)| \]

The proof of ii) is analogous.

The following theorem will show that \( \Delta^n(\cdot) \) is uniformly continuous over the real line.

Theorem 2.6.2:

\[ |\Delta^n(\delta + h, I_1, I_2) - \Delta^n(\delta, I_1, I_2)| < 2^{n-1}|h|, \text{ for all } I_1, I_2, \delta, h. \]

Proof: Let \( n=1 \), then

\[ |\Delta^1(\delta + h, I_1, I_2) - \Delta^1(\delta, I_1, I_2)| = |\delta + h - \delta| = |h| \]

Suppose the theorem holds for \( n \); we need to show that it holds for \( n+1 \).
\[ |\Delta^{n+1}(\delta+h, I_1, I_2) - \Delta^{n+1}(\delta, I_1, I_2) | \leq \int |\Delta^{n}(\delta + h - z(I_1(I_1+1))^{-1/2}, I_1+1, I_2)| \\
- \Delta^{n}(\delta - z(I_1(I_1+1))^{-1/2}, I_1+1, I_2) |d\phi(z) \\
+ \int |\Delta^{n}(\delta + h + z(I_2(I_2+1))^{-1/2}, I_1, I_2+1)| \\
- \Delta^{n}(\delta + z(I_2(I_2+1))^{-1/2}, I_1, I_2+1) |d\phi(z) \\
\leq 2^{n-1} |h| + 2^{n-1} |h| \\
= 2^n |h| \]

The first inequality results from standard absolute value inequalities while the second inequality is a direct consequence of Lemma 2.6.1. The induction hypothesis implies the third inequality.

Q.E.D.
Corollary 2.6.3:

\[ \Delta^n(\delta, I_1, I_2) \] is a uniformly continuous function of \( \delta \), for all fixed values of \( I_1, I_2 \).

Notice that we really have a stronger result than Corollary 2.6.3, since the bound does not depend on the information. The technical term for such independence is equicontinuity of \( \Delta^n(\cdot) \) with respect to \( I_1 \) and \( I_2 \).

The next theorem will give a bound on \( \Delta^n(\delta, I_1, I_2) \) that will be used to justify the use of Lebesgue's Dominated Convergence Theorem in Theorem 2.6.5.

Theorem 2.6.4:

\[ |\Delta^n(\delta, I_1, I_2)| \leq 2^{n-1} |\delta| + \left( \sum_{i=1}^n (2^{n-1} - 1) \left( \frac{1}{I_1(I_1+1)} \right)^{-1/2} + \frac{1}{I_2(I_2+1)} \right)^{1/2} \]

Proof:

Let \( n=1 \), then

\[ |\Delta^1(\delta, I_1, I_2)| = \delta. \]

Suppose the theorem holds for \( n \); we need to show that it holds for \( n+1 \).
\[ |\Delta^{n+1}(\delta, I_1, I_2)| = \left| \int \Delta^n (\delta - z(I_1 + 1))^{-1/2}, I_1 + 1, I_2) d\phi(z) \right| \]

\[ + \int \Delta^n (\delta + z(I_2 + 1))^{-1/2}, I_1, I_2 + 1) d\phi(z) \]

\[ \leq \int |\Delta^n (\delta - z(I_1 + 1))^{-1/2}, I_1 + 1, I_2) d\phi(z) | \]

\[ + \int |\Delta^n (\delta + z(I_2 + 1))^{-1/2}, I_1, I_2 + 1) d\phi(z) | \]

\[ \leq \int |\Delta^n (\delta - z(I_1 + 1))^{-1/2}, I_1 + 1, I_2) d\phi(z) | \]

\[ + \int |\Delta^n (\delta + z(I_2 + 1))^{-1/2}, I_1, I_2 + 1) d\phi(z) | \]

\[ = 2^n |\delta| + \frac{2}{\pi} (2^n - 2) ((I_1(I_1 + 1))^{-1/2} + (I_2(I_2 + 1))^{-1/2}) \int |z| d\phi(z) \]

Equation 2.4.12 implies the first equality while the next inequality follows from absolute value inequalities. The final inequality holds since \(|f(x)^+| < |f(x)|\) and \(|f(x)^-| < |f(x)|\).

Now, if we apply the induction hypothesis to Equation 2.6.1 we have that

\[ |\Delta^{n+1}(\delta, I_1, I_2)| \leq \int 2^{n-1} (|\delta - z(I_1 + 1)|^{-1/2}) \]

\[ + (\frac{2}{\pi} (2^n - 2)) ((I_1(I_1 + 1))^{-1/2} + (I_2(I_2 + 1))^{-1/2}) \int |z| d\phi(z) \]

\[ \leq 2^n |\delta| + ((I_1(I_1 + 1))^{-1/2} + (I_2(I_2 + 1))^{-1/2}) \int |z| d\phi(z) \]

\[ + (\frac{2}{\pi} (2^n - 2)) ((I_1(I_1 + 1))^{-1/2} + (I_2(I_2 + 1))^{-1/2}) \]

\[ = 2^n |\delta| + \frac{2}{\pi} (2^n - 1) ((I_1(I_1 + 1))^{-1/2} + (I_2(I_2 + 1))^{-1/2}) . \]
The last inequality and equality follow from simplification
of Equation 2.6.1.
Q.E.D.

The following theorem provides a limiting rate of change in
\( \Delta^n(\cdot) \) and is a direct extension of part ii of Lemma 2.5.1.

**Theorem 2.6.5:**

i) \[
\lim_{\delta \to \infty} \frac{\Delta^n(\delta, I_1, I_2)}{\delta} = 1
\]

ii) \[
\lim_{\delta \to \infty} \frac{\Delta^n(\delta, I_1, I_2)}{\delta} = 1
\]

**Proof of i:** Let \( n=1 \), then

\[
\Delta^1(\delta, I_1, I_2) = 1.
\]

Suppose the theorem holds for \( n \); we need to show that it
holds for \( n+1 \).

\[
\lim_{\delta \to \infty} \frac{\Delta^{n+1}(\delta, I_1, I_2)}{\delta} = \lim_{\delta \to \infty} \left( \frac{\Delta^n(\delta-z(I_1(I_1+1))^{-1/2}, I_1+1, I_2)}{\delta} \right)^+ + \lim_{\delta \to \infty} \left( \frac{\Delta^n(\delta+z(I_2(I_2+1))^{-1/2}, I_1, I_2+1)}{\delta} \right)^-
\]

Now if we use Lebesgue's Dominated Convergence Theorem twice

where the dominating functions are

\[
2^{n-1} |\delta-z(I_1(I_1+1))^{-1/2}| + \left( \frac{1}{\sqrt{n}} \right) (2^{n-1}-1) ((I_1(I_1+1))^{-1/2} + \frac{1}{2}) \delta^{-1}
\]

and

\[
2^{n-1} |\delta+z(I_2(I_2+1))^{-1/2}| + \left( \frac{1}{\sqrt{n}} \right) (2^{n-1}-1) ((I_1(I_1+1))^{-1/2} + \frac{1}{2}) \delta^{-1}
\]
we have the following equality

\[
\lim_{\delta \to +\infty} \Delta^{n+1}(\delta, I_1, I_2) = \lim_{\delta \to +\infty} \Delta^n(\delta-z(I_1(I_1+1)^{-1/2}, I_1+1, I_2) + \\
\lim_{\delta \to +\infty} \Delta^n(\delta+z(I_2(I_2+1)^{-1/2}, I_1, I_2) - d\phi(z)
\]

Now,

\[
\lim_{\delta \to +\infty} \frac{\delta-z(I_1(I_1+1))^{-1/2}}{\delta} = \lim_{\delta \to +\infty} \frac{\delta-z(I_1(I_1+1))^{-1/2}}{\delta}
\]

\[
\times \left[ \frac{1}{\delta} \right] = 1, \text{ for all } z.
\]

using the induction hypothesis.

Therefore,

\[
\lim_{\delta \to +\infty} \frac{\delta-z(I_1(I_1+1))^{-1/2}}{\delta} = 1.
\]

since Equation 2.6.4 implies that \(\Delta^n(\delta-z(I_1(I_1+1))^{-1/2}, I_1+1, I_2) > 0\) for large \(\delta\).

Similarly, the induction hypothesis gives that

\[
\lim_{\delta \to +\infty} \frac{\delta-z(I_2(I_2+1))^{-1/2}}{\delta} = 1
\]

and

\[
\lim_{\delta \to +\infty} \frac{\delta-z(I_2(I_2+1))^{-1/2}}{\delta} = 0.
\]
Therefore, if we combine Equations 2.6.3-2.6.5 we have that

\[ \lim_{\delta \to \infty} \frac{\Delta^{n+1}(\delta, I_1, I_2)}{\delta} = \int 0d\Phi(z) + \int 1d\Phi(z) \]

= 1.

Proof of ii:

Reasoning identical to that producing 2.6.3-2.6.5 provides the following three equalities.

\[ \lim_{\delta \to \infty} \frac{\Delta^{n+1}(\delta, I_1, I_2)}{\delta} = \int \lim_{\delta \to \infty} \frac{\Delta^n(\delta-z(I_1(I_1+1))^{-1/2}, I_1+1, I_2)}{\delta} d\Phi(z) \]

\[ + \int \lim_{\delta \to \infty} \frac{\Delta^n(\delta+z(I_2(I_2+1))^{-1/2}, I_1, I_2+1)}{\delta} d\Phi(z) \]

(2.6.6)

and

\[ \lim_{\delta \to \infty} \frac{\Delta^n(\delta-z(I_1(I_1+1))^{-1/2}, I_1+1, I_2)}{\delta} = 0 \]  

(2.6.7)

and

\[ \lim_{\delta \to \infty} \frac{\Delta^n(\delta+z(I_2(I_2+1))^{-1/2}, I_1, I_2+1)}{\delta} = 1. \]  

(2.6.8)

Therefore, if we combine 2.6.6-2.6.8 we have that

\[ \lim_{\delta \to \infty} \frac{\Delta^n(\delta, I_1, I_2)}{\delta} = \int 0d\Phi(z) + \int 1d\Phi(z) \]

= 1

Q.E.D.
The following theorem will give the structure of $\Delta^n(\cdot)$ for large informations and is a direct extension of part i of Lemma 2.5.1.

**Theorem 2.6.6:**

$$\lim_{\min(I_1, I_2) \to \infty} \Delta^n(\delta, I_1, I_2) = \delta.$$  

**Proof:** Let $n=1$, then

$$\Delta^1(\delta, I_1, I_2) = \delta.$$  

In the remainder of this proof, we will suppress the subscript in the limit notation.

$$\lim \Delta^{n+1}(\delta, I_1, I_2) = \lim \Delta^n(\delta - z(I_1(I_1+1))^{-1/2}, I_1+1, I_2) + \Phi(z) + \lim \Delta^n(\delta + z(I_2(I_2+1))^{-1/2}, I_1, I_2+1) \Phi(z).$$

(2.6.9)

Now, if we use Lebesgue's Dominated Convergence Theorem twice where the dominating functions are

$$2^{n-1} |\delta - z| + \left(\frac{\delta}{n}\right) (2^{n-1}-1),$$  

$$2^{n-1} |\delta + z| + \left(\frac{\delta}{n}\right) (2^{n-1}-1),$$

we have the following equality

$$\lim \Delta^{n+1}(\delta, I_1, I_2) = \lim \Delta^n(\delta - z(I_1(I_1+1))^{-1/2}, I_1+1, I_2) + \Phi(z) + \lim \Delta^n(\delta + z(I_2(I_2+1))^{-1/2}, I_1, I_2+1) \Phi(z).$$

Therefore, if we could show that
\[ \lim \Delta^n(\delta-z(I_{1}(I_{1}+1))^{1/2},I_{1}+1,I_{2}) = \delta \] (2.6.10)

and

\[ \lim \Delta^n(\delta+z(I_{2}(I_{2}+1))^{1/2},I_{2}+1,I_{2}) = \delta \] (2.6.11)

we would be done, since Equation 2.6.9 would simplify to

\[ \lim \Delta^{n+1}(\delta,I_{1},I_{2}) = \int \max(0,\delta)d\Phi(z) + \int \min(0,\delta)d\Phi(z) = \delta. \]

The following argument proves Equation 2.6.10. Let \( \varepsilon > 0 \) be given; then

\[ |\Delta^n(\delta-z(I_{1}(I_{1}+1))^{1/2},I_{1}+1,I_{2}) - \delta| \leq |\Delta^n(\delta-z(I_{1}(I_{1}+1))^{1/2},I_{1}+1,I_{2}) - \delta|
\]

\[ - \Delta^n(\delta,I_{1}+1,I_{2}) | + |\Delta^n(\delta,I_{1}+1,I_{2}) - \delta| \]

(2.6.12)

Now,

i) \[ |\Delta^n(\delta-z(I_{1}(I_{1}+1))^{1/2},I_{1}+1,I_{2})
\]

\[ - \Delta^n(\delta,I_{1}+1,I_{2}) | \leq 2^{n-1}|z|(I_{1}(I_{1}+1))^{1/2} \]

\[ \leq \varepsilon/2, \text{ for all } I_{1} > k_{1}(\varepsilon,z) \]

The first inequality follows from Theorem 2.6.2.

ii) \[ |\Delta^n(\delta,I_{1}+1,I_{2}) - \delta| < \varepsilon/2, \text{ whenever } I_{1},I_{2} > k_{2}(\varepsilon,\delta), \]

using the induction hypotheses.

Equation 2.6.10 follows directly from i, ii and Equation 2.6.12.

Reasoning identical to the above shows that Equation 2.6.11 holds.

Q.E.D.
Intuitively, Theorem 2.6.6 means that for large informations the optimal strategy is entirely determined by the current difference in the posterior means.

The next theorem will show that \( \Delta^n(\cdot) \) is an increasing function of the difference of the means \( \mu_2 - \mu_1 \).

**Theorem 2.6.7:** \( \Delta^n(\delta, I_1, I_2) \) is an increasing function of \( \delta \), for every fixed value of \( I_1 \) and \( I_2 \).

**Proof:** Let \( n=1 \), then \( \Delta^1(\delta, I_1, I_2) = \delta \) which is obviously an increasing function of \( \delta \).

Let the induction hypothesis be true for \( n \); we need to show that it is true for \( n+1 \).

Assume that \( \delta' < \delta'' \); therefore, we must show that

\[
\Delta^{n+1}(\delta'', I_1, I_2) - \Delta^{n+1}(\delta', I_1, I_2) > 0.
\]

Equation 2.4.12 implies that

\[
\Delta^{n+1}(\delta'', I_1, I_2) - \Delta^{n+1}(\delta', I_1, I_2) = \left( \frac{\Delta^n(\delta'' - z(I_1(I_1 + 1)))^{-1/2}, I_1 + 1, I_2)}{\Delta^n(\delta' - z(I_1(I_1 + 1)))^{-1/2}, I_1 + 1, I_2)}\right) d\Phi(z) \\
+ \int \left( \frac{\Delta^n(\delta'' + z(I_2(I_2 + 1)))^{-1/2}, I_1, I_2 + 1)}{\Delta^n(\delta' - z(I_2(I_2 + 1)))^{-1/2}, I_1, I_2 + 1)}\right) d\Phi(z)
\]

(2.6.13)
But, by the induction hypothesis we have that

$$\Delta^n(\delta'' - z(I_1(I_1+1))^{-1/2}, I_1+1, I_2) - \Delta^n(\delta' - z(I_1(I_1+1))^{-1/2}, I_1+1, I_2) > 0$$

and

$$\Delta^n(\delta' + z(I_2(I_2+1))^{-1/2}, I_1, I_2+1) - \Delta^n(\delta' - z(I_2(I_2+1))^{-1/2}, I_1, I_2+1) > 0,$$

for all $z$.

which implies that

$$\Delta^n(\delta'' - z(I_1(I_1+1))^{-1/2}, I_1+1, I_2) + \Delta^n(\delta' - z(I_1(I_1+1))^{-1/2}, I_1+1, I_2) > 0$$

(2.6.14)

and

$$\Delta^n(\delta' + z(I_2(I_2+1))^{-1/2}, I_1, I_2+1) - \Delta^n(\delta' - z(I_2(I_2+1))^{-1/2}, I_1, I_2+1) > 0,$$

for all $z$. (2.6.15)

The following argument is needed to ensure that

$$\Delta^{n+1}(\delta'', I_1, I_2) - \Delta^n(\delta', I_1, I_2)$$

is strictly greater than zero.

By Theorem 2.6.5 we know that there exists a $\delta^*$ such that

$$\Delta^n(\delta, I_1, I_2) > 0, \text{ for all } \delta > \delta^*.$$

Therefore, there exists a $z^*$ such that

$$\Delta^n(\delta'' - z(I_1(I_1+1))^{-1/2}, I_1+1, I_2) + \Delta^n(\delta' - z(I_2(I_2+1))^{-1/2}, I_1+1, I_2) > 0,$$

for all $z < z^*$. (2.6.16)

Now, if we combine 2.6.13-2.6.16, we have that
\[ \int [\Delta^n(\delta''-z(I_1(I_2+1))^{-1/2},I_1+1,I_2^+)]d\phi(z) > 0 \]

and

\[ \int [\Delta^n(\delta''+z(I_2(I_2+1))^{-1/2},I_1^+,I_2^+)]d\phi(z) > 0 \]

which implies that

\[ \Delta^{n+1}(\delta'',I_1,I_2) - \Delta^{n+1}(\delta',I_1,I_2) > 0. \]

**Corollary 2.6.8:** There exists a unique number \( \delta^* \) such that

i) \( \Delta^n(\delta^*,I_1,I_2) = 0 \)

ii) \( \Delta^n(\delta,I_1,I_2) \leq 0, \) if \( \delta \leq \delta^* \)

**Proof of i:**

Theorem 2.6.5 implies that there exist two points, \( \delta' \) and \( \delta'' \), such that

i) \( \Delta^n(\delta',I_1,I_2) < 0 \)

ii) \( \Delta^n(\delta'',I_1,I_2) > 0. \)

Now, since \( \Delta^n(\delta,I_1,I_2) \) is a continuous function of \( \delta \), we know that there must exist at least one point, \( \delta^* \), such that

\[ \Delta^n(\delta^*,I_1,I_2) = 0. \]

Now, suppose there were two such points, say \( \delta^*_1 \) and \( \delta^*_2 \), then we would have that
i) $\Delta^n(\delta^*_{1}, I_{1}, I_{2}) = \Delta^n(\delta^*_{2}, I_{1}, I_{2})$

ii) $\delta^*_{1} \neq \delta^*_{2}$.

which is a contradiction to the strictly increasing nature of $\Delta^n(\cdot)$ with respect to $\delta$. Therefore, $\delta^*$ is unique and has the desired property.

Proof of ii: By i) we know that there exists a unique $\delta$ such that $\Delta^n(\delta, I_{1}, I_{2}) = 0$. Therefore, since $\Delta^n(\cdot)$ is a strictly increasing function of $\delta$ we have the desired result.

Corollary 2.6.8 is significant, since the optimal strategy can be characterized as follows:

i) sample source $X$ if $\delta > \delta^n(I^*_1, I^*_2)$

ii) sample source $Y$ if $\delta < \delta^n(I^*_1, I^*_2)$

iii) sample either source if $\delta = \delta^n(I^*_1, I^*_2)$

where

$n =$ the number of observations still to be taken

$\delta =$ current difference in the posterior means

$I^*_1 =$ current information on $X$

$I^*_2 =$ current information on $Y$

$\delta^n(I^*_1, I^*_2)$ is the root of $\Delta^n(\delta, I_{1}, I_{2})$.

Unlike many of the quantities that have been discussed, $\delta^n(I^*_1, I^*_2)$
does not have a recursive structure. Therefore, we must actually solve for the root in order to find \( \delta^n(I_1, I_2) \), but Lemma 2.6.9 implies that we need only calculate the root for situations where \( I_1 > I_2 \).

**Lemma 2.6.9**: \( \delta^n(I_1, I_2) = -\delta^n(I_2, I_1) \)

**Proof**: \( \Delta^n(-\delta^n(I_2, I_1), I_1, I_2) = -\Delta^n(\delta^n(I_2, I_1), I_2, I_1) \)

\[ = 0. \]

The above two equalities are implied by Lemma 2.4.2 and the fact that \( \delta^n(I_2, I_1) \) is a root of \( \Delta^n(\delta, I_2, I_1) \), respectively. Therefore, \( -\delta^n(I_2, I_1) \) is a root of \( \Delta^n(\delta, I_1, I_2) \), but the root of \( \Delta^n(\delta, I_1, I_2) \) is unique. Hence, \( \delta^n(I_1, I_2) = -\delta^n(I_2, I_1) \).

2.7. Bound on \( \Delta^n(\cdot) \)

This section is devoted to finding a tight bound for \( \Delta^n(\cdot) \) when the difference in the information on the sources is greater than or equal to \( n-2 \). First, we will develop the bound for the case where \( n=2 \) by maximizing

\[ \Delta^2(\delta, I_1, I_2) - \delta \]

with respect to \( \delta \). By exploiting the similarity of the case of arbitrary \( n \) to the case where \( n=2 \) we will extend to larger \( n \).
Let $g_2(\delta, I_1, I_2) \equiv \Delta^2(\delta, I_1, I_2) - \delta$

then

$$g_2(\delta, I_1, I_2) = \delta\phi(\delta(I_1(I_1+1))^{1/2}) + \phi(-\delta(I_2(I_2+1))^{1/2}) - 1$$

$$+ (I_1(I_1+1))^{-1/2}\phi(\delta(I_1(I_1+1))^{1/2})$$

$$- (I_2(I_2+1))^{-1/2}\phi(-\delta(I_2(I_2+1))^{1/2})$$

(2.7.1)

using Equation 2.5.1.

Now,

$$\frac{d}{d\delta} g_2(\delta, I_1, I_2) = \frac{d}{d\delta} \Delta^2(\delta, I_1, I_2) - 1$$

$$= \phi(\delta(I_1(I_1+1))^{1/2}) + \phi(-\delta(I_2(I_2+1))^{1/2}) - 1$$

$$= \phi(\delta(I_1(I_1+1))^{1/2}) - \phi(\delta(I_2(I_2+1))^{1/2}).$$

(2.7.2)

The last two equalities follow from Equation 2.5.2 and the symmetry of the standard normal distribution.

First, we will consider the situation where $I_2 \geq I_1$. Therefore, Equation 2.7.2 can be rewritten as

$$\frac{d}{d\delta} g_2(\delta, I_1, I_2) = \begin{cases} 
-\Pr(\delta(I_1(I_1+1))^{1/2} < z < \delta(I_2(I_2+1))^{1/2}) & \text{if } \delta > 0 \\
\Pr(-\delta(I_1(I_1+1))^{1/2} < z < -\delta(I_2(I_2+1))^{1/2}) & \text{if } \delta < 0
\end{cases}$$

$$= -\text{sgn}(\delta)\Pr(|\delta|(I_1(I_1+1))^{1/2} < z|\delta|(I_2(I_2+1))^{1/2})$$

(2.7.3)
where $z$ is a standard normal random variable.

Equation 2.7.3 will be used to prove Lemma 2.7.1.

**Lemma 2.7.1:**

i) $\max g_2(\delta, I_1, I_2) = g_2(0, I_1, I_2)$

ii) $g_2(\delta, I_1, I_2) \geq 0$

**Proof of i:**

Equation 2.7.3 implies that $g_2(\delta, I_1, I_2)$ is a strictly decreasing function of $\delta$ if $\delta > 0$ and a strictly increasing function of $\delta$ if $\delta < 0$. Therefore, $g_2(\delta, I_1, I_2)$ is maximized when $\delta = 0$.

**Proof of ii:**

Since $g_2(\delta, I_1, I_2)$ is a strictly decreasing function of $\delta$ if $\delta < 0$, we have that

$$g_2(\delta, I_1, I_2) \geq \lim_{\delta \to \infty} g_2(\delta, I_1, I_2)$$

$$= \lim_{\delta \to \infty} \delta [\Phi(\delta (I_1 (I_1+1)^{1/2}) + \Phi(-\delta (I_2 (I_2+1)^{1/2}) - 1]

\geq \lim_{\delta \to \infty} -\delta [1-\Phi(\delta (I_1 (I_1+1)^{1/2})]

= -(I_1 (I_1+1))^{-1/2} \lim_{\delta \to \infty} \delta (I_1 (I_1+1))^{1/2} [1-\Phi(\delta (I_1 (I_1+1))^{1/2})

= -(I_1 (I_1+1))^{-1/2} \lim_{\delta \to \infty} \Phi(\delta (I_1 (I_1+1))^{1/2})

= 0.$$
The first equality follows from taking the limit of Equation 2.7.1. The second inequality and equality follow from simplification and rearrangement. The third equality is a direct consequence of Equation 1.8 of Feller (1957). Therefore, \( g_2(\delta, I_1, I_2) > 0 \) for all \( \delta > 0 \).

An identical argument produces the fact that \( g_2(\delta, I_1, I_2) \geq 0 \) for all \( \delta < 0 \).

Q.E.D.

Lemma 2.7.1 will be used to prove the following lemma.

**Lemma 2.7.2:**
\[
\Delta^2(\delta, I_1, I_2) < \delta + \frac{1}{\sqrt{2\pi}} \left( (I_1(I_1+1))^{-1/2} - (I_2(I_2+1))^{-1/2} \right) \text{ if } I_2 \geq I_1
\]

**Proof:**
\[
\Delta^2(\delta, I_1, I_2) = \delta + \Delta^2(\delta, I_1, I_2) - \delta = \delta + g_2(\delta, I_1, I_2)
\]
\[
\leq \delta + \max_{\delta} g_2(\delta, I_1, I_2) = \delta + g_2(0, I_1, I_2)
\]
\[
= \delta + \frac{1}{\sqrt{2\pi}} \left( (I_1(I_1+1))^{-1/2} - (I_2(I_2+1))^{-1/2} \right).
\]

The second and fourth equalities follow from the definition of \( g_2(\delta, I_1, I_2) \) while the third equality follows from Lemma 2.7.1, part i.

Q.E.D.

Now, we are in a position to extend the bound to larger values of \( n \).
Theorem 2.7.3:

\[ \Delta^n(\delta, I_1, I_2) \leq \delta + h_n(I_1, I_2), \text{ if } I_2 - I_1 \geq n-2 \]

where

\[ h_n(I_1, I_2) = \frac{1}{\sqrt{2\pi}} \left( (n-1)(I_1(I_1+1))^{-1/2} - \sum_{i=1}^{n-1} \left( (I_2+i-1)(I_2+i) \right)^{-1/2} \right) \]

Proof:

Notice that the bound has been proven to hold for the case \( n=2 \) in Lemma 2.7.2.

Suppose the bound holds for \( n \); we need to show that it holds for \( n+1 \). Now, Equation 2.4.12 implies that

\[ \Delta^{n+1}(\delta, I_1, I_2) = \int \Delta^n(\delta-z(I_1(I_1+1))^{-1/2}, I_1+1, I_2) \, d\phi(z) \]

\[ + \int \Delta^n(\delta+z(I_2(I_2+1))^{-1/2}, I_1, I_2+1) \, d\phi(z). \] (2.7.4)

But, \( I_2 - I_1 \geq (n+1)-2 \) implies that both integrands of 2.7.4 may be bounded.

Therefore,

\[ \Delta^{n+1}(\delta, I_1, I_2) \leq \int [\delta-z(I_1(I_1+1))^{-1/2} + h_n(I_1+1, I_2)] \, d\phi(z) \]

\[ + \int [\delta+z(I_2(I_2+1))^{-1/2} + h_n(I_1, I_2+1)] \, d\phi(z) \]

\[ \leq \int [\delta+h_n(I_1, I_2+1) - z(I_1(I_1+1))^{-1/2}] \, d\phi(z) \]

\[ + \int [\delta+h_n(I_1, I_2+1) + z(I_2(I_2+1))^{-1/2}] \, d\phi(z) \]

\[ = \Delta^2(\delta+h_n(I_1, I_2+1), I_1, I_2). \] (2.7.5)
The first inequality results from using the bound on both integrands of 2.7.4. The next inequality follows from the fact that
\[ h_n(I_1 + 1, I_2) < h_n(I_1, I_2 + 1). \]

The last equality is a restatement of Equation 2.4.12.

Now, we are in a position to use the same development as in part i of Lemma 2.7.1.

Define
\[ g_{n+1}(\delta, I_1, I_2) = \Delta^2(\delta + h_n(I_1, I_2 + 1), I_1, I_2) - \delta. \]

Therefore,
\[
\frac{d}{d\delta} g_{n+1}(\delta, I_1, I_2) = \frac{d}{d\delta} g_2(\delta + h_n(I_1, I_2 + 1), I_1, I_2)
= -\text{sgn}(\delta + h_n(I_1, I_2 + 1))* \Pr(|\delta + h_n(I_1, I_2 + 1)|(I_1(I_2 + 1))^{1/2} < z)
\leq |\delta + h_n(I_1, I_2 + 1)|(I_2(I_2 + 1))^{1/2}. \quad (2.7.6)
\]

The first equality is an application of the chain rule of calculus while the second is a restatement of Equation 2.7.3.

Equation 2.7.6 implies that \( g_{n+1}(\delta, I_1, I_2) \) is a strictly decreasing function of \( \delta \) if \( \delta > -h_n(I_1, I_2 + 1) \) and a strictly increasing function of \( \delta \) if \( \delta < -h_n(I_1, I_2 + 1) \). Therefore, \( g_{n+1}(\delta, I_1, I_2) \) is maximized when \( \delta = -h_n(I_1, I_2 + 1) \).
Hence,

\[
\max_\delta g_{n+1}(\delta, I_1, I_2) = g_{n+1}(-h_n(I_1, I_2+1), I_1, I_2)
\]

\[
= \Delta^2(-h_n(I_1, I_2+1) + h_n(I_1, I_2+1), I_1, I_2)
\]

\[
= \Delta^2(0, I_1, I_2) + h_n(I_1, I_2+1)
\]

\[
= \frac{1}{\sqrt{2\pi}} \left[ \left((I_1(I_1+1)) - \frac{1}{2} - (I_2(I_2+1)) - \frac{1}{2} \right) \right.
\]

\[
+ \frac{1}{\sqrt{2\pi}} \left[ \left((n-I_1(I_1+1)) - \frac{1}{2} - \sum_{i=1}^{n-1} ((I_2+i)(I_2+i+1)) - \frac{1}{2} \right) \right]
\]

\[
= \frac{1}{\sqrt{2\pi}} \left[ n(I_1(I_1+1)) - \frac{n}{2} - \sum_{i=1}^{n} ((I_2+i)(I_2+i+1)) - \frac{1}{2} \right]
\]

\[
= h_{n+1}(I_1, I_2). \tag{2.7.7}
\]

The first equality follows from the fact that the maximum of \(g_{n+1}(\delta, I_1, I_2)\) occurs at \(-h_n(I_1, I_2+1)\) while the second follows from definition of \(g_n(\delta, I_1, I_2)\). The last three equalities result from definition of \(h_n(I_1, I_2)\).

Now,

\[
\Delta^{n+1}(\delta, I_1, I_2) \leq \Delta^2(\delta + h_n(I_1, I_2+1), I_1, I_2)
\]

\[
= \delta + (\Delta^2(\delta + h_n(I_1, I_2+1), I_1, I_2) - \delta)
\]

\[
= \delta + g_{n+1}(\delta, I_1, I_2)
\]

\[
\leq \delta + \max_\delta g_{n+1}(\delta, I_1, I_2)
\]

\[
= \delta + h_{n+1}(I_1, I_2).
\]
The first inequality is a restatement of Equation 2.7.5 while the second equality uses the definition of $g_{n+1}(\delta, I_1, I_2)$. The last equality is a restatement of Equation 2.7.7.

Therefore, we have shown that the bound holds for $n+1$.

Q.E.D.

The following theorem is a direct extension of part ii of Lemma 2.7.1.

Theorem 2.7.4: $g_{n+1}(\delta, I_1, I_2) \geq 0$ if $I_2 > I_1$

Proof:

\[
g_{n+1}(\delta, I_1, I_2) = \Delta^2(\delta + h_n(I_1, I_2+1), I_1, I_2) - \delta
\]

\[
= \Delta^2(\delta + h_n(I_1, I_2+1), I_1, I_2) - (\delta + h_n(I_1, I_2+1)) + h_n(I_1, I_2+1)
\]

\[
\geq h_n(I_1, I_2+1)
\]

\[
\geq 0
\]

The first equality and the third equality are restatements of the definition $g_{n+1}(\delta, I_1, I_2)$ while the first inequality is a consequence of Lemma 2.7.1, part ii. The final inequality follows from the definition of $h_n(I_1, I_2)$.

Q.E.D.
So, in summary, we have proven the following properties. If $I_2 - I_1 \geq n-2$, then

i) $\max_\delta q_n(\delta, I_1, I_2) = q_n(-h_n(I_1, I_2 + 1), I_1, I_2) = h_{n+1}(I_1, I_2)$

ii) $\Delta_n^+(\delta, I_1, I_2) \leq \delta + q_n(\delta, I_1, I_2)$

iii) $q_n(\delta, I_1, I_2) \geq 0$.

Now, we are in a position to graphically show how the functions $\delta$, $\delta + q_n(\delta, I_1, I_2)$, and $\delta + h_{n+1}(I_1, I_2)$ are related.

Up until now, we have only considered the situation where the objective was providing an upper bound for $\Delta^+(\delta, I_1, I_2)$ when $I_2 - I_1 > n-2$. In the following corollary we will establish a lower bound for $\Delta^+(\delta, I_1, I_2)$ when $I_1 - I_2 > n-2$.

**Corollary 2.75:** $\Delta^+(\delta, I_1, I_2) \geq \delta - h_n(I_2, I_1)$ if $I_1 - I_2 \geq n-2$

**Proof:**

$I_1 - I_2 \geq n-2 \Rightarrow \Delta^+(\delta, I_2, I_1) - \delta < h_n(I_2, I_1)$, for all $\delta$

$\Rightarrow -\Delta^-(\delta, I_1, I_2) - \delta < h_n(I_2, I_1)$, for all $\delta$

$\Rightarrow -\Delta^+(\delta^*, I_1, I_2) + \delta^* < h_n(I_2, I_1)$, for all $\delta^*$

$\Rightarrow \Delta^+(\delta^*, I_1, I_2) > \delta - h_n(I_2, I_1)$, for all $\delta^*$

The first implication is a restatement of Theorem 2.7.3 where the roles of $I_1$ and $I_2$ have been interchanged. The second implication is an application of Theorem 2.4.2 while the third...
Figure 2.7.1. Graphical presentation of Theorem 2.7.3
implication follows from substituting \( \delta^* \) for \( -\delta \).

Q.E.D.

The following corollary provides an upper bound for the root of \( \Delta^n(\cdot) \).

\textbf{Corollary 2.7.6:} \( \delta^n(I_1, I_2) \leq h_n(I_2, I_1) \) if \( I_1 - I_2 \geq n-2 \)

where \( \delta^n(I_1, I_2) \) is the root of \( \Delta^n(\delta, I_1, I_2) \).

\textbf{Proof:} Notice that

\[
0 = \Delta^n(\delta^n(I_1, I_2), I_1, I_2)
\]

\[
\geq \delta^n(I_1, I_2) - h_n(I_2, I_1).
\]

\[
\delta^n(I_1, I_2) \leq h_n(I_2, I_1).
\]

\textbf{2.8. Approximation of } \( \Delta^n(\cdot) \)

In this section, we develop three approximations of \( \Delta^n(\cdot) \). All the approximations are linear functions of \( \delta \) where the slope and intercept are determined by the informations. Therefore, the approximations have the structure, \( a_n(I_1, I_2)\delta + b_n(I_1, I_2) \). The choices of \( a_n(I_1, I_2) \) and \( b_n(I_1, I_2) \) are based on repeated first order Taylor series expansions. Also, the slope and intercept for every value of \( n \) possess a recursive structure similar to that of \( \Delta^n(\cdot) \). Justification for each approximation is given in several related theorems.

We begin by investigating approximations to \( \Delta^2(\delta, I_1, I_2) \) and \( \Delta^3(\delta, I_1, I_2) \). From Equations 2.5.1 and 2.5.2 we have that
\[ \Delta^2(\delta, I_1, I_2) = \delta(\delta(I_1(I_1+1))^{1/2}) + (-\delta(I_2(I_2+1))^{1/2}) \]
\[ + (I_1(I_1+1))^{-1/2}\delta(I_1(I_1+1))^{1/2} \]
\[ - (I_2(I_2+1))^{-1/2}\delta(I_2(I_2+1))^{1/2} \]

and
\[ \frac{d}{d\delta} \Delta^2(\delta, I_1, I_2) = \delta(\delta(I_1(I_1+1))^{1/2}) + (-\delta(I_2(I_2+1))^{1/2}). \]

One approximation involves the first order Taylor series expansion of \( \Delta^2(\delta, I_1, I_2) \) about zero given by
\[ \Delta^2_*(\delta, I_1, I_2) = \frac{1}{\sqrt{2\pi}} \int \Delta^2(\delta, I_1, I_2) d\phi(z) \]
\[ = \delta + \frac{1}{\sqrt{2\pi}}((I_1(I_1+1))^{-1/2} - (I_2(I_2+1))^{-1/2}). \] (2.8.1a)

Recall from Equation 2.4.12 that
\[ \Delta^3(\delta, I_1, I_2) = \int \Delta^2(\delta-z(I_1(I_1+1))^{-1/2} - (I_2(I_2+1))^{-1/2}) d\phi(z) \]
\[ + \int \Delta^2(\delta+z(I_2(I_2+1))^{-1/2} - (I_2(I_2+1))^{-1/2}) d\phi(z). \] (2.8.1b)

An approximation of \( \Delta^3(\delta, I_1, I_2) \) could be obtained by substituting \( \Delta^2_*(\delta, I_1, I_2) \) for \( \Delta^2(\delta, I_1, I_2) \) in Equation (2.8.1b)

After substitution, we have that
\[ \Delta^3_*(\delta, I_1, I_2) = \int \Delta^2(\delta-z(I_1(I_1+1))^{-1/2} + b_2(I_2(I_2+1))^{-1/2}) d\phi(z) \]
\[ + \int \Delta^2(\delta+z(I_2(I_2+1))^{-1/2} + b_2(I_2(I_2+1))^{-1/2}) d\phi(z) \]

where
\[ b_2(I_1, I_2) = \frac{1}{\sqrt{2\pi}}((I_1(I_1+1))^{-1/2} - (I_2(I_2+1))^{-1/2}). \]

The RHS of the above equation will be denoted by \( \Delta^3_*(\delta, I_1, I_2) \).
\[ \Delta_{**}^3(\delta, I_1, I_2) = (\delta+b_2(I_1+1, I_2))\Phi((\delta+b_2(I_1+1, I_2))(I_1(I_1+1))^{1/2}) \\
+ (\delta+b_2(I_1, I_2+1))\Phi(-(\delta+b_2(I_1, I_2+1))(I_2(I_2+1))^{1/2}) \\
+ (I_1(I_1+1))^{-1/2}\Phi((-\delta+b_2(I_1, I_2+1))(I_1(I_1+1))^{1/2}) \\
- (I_1(I_1+1))^{-1/2}\Phi((\delta+b_2(I_1, I_2+1))(I_2(I_2+1))^{1/2}) \]

and

\[ \frac{d}{d\delta} \Delta_{**}^3(\delta, I_1, I_2) = \Phi((\delta+b_2(I_1+1, I_2))(I_1(I_1+1))^{1/2}) \\
+ \Phi(-\delta+b_2(I_1, I_2+1))(I_2(I_2+1))^{1/2}) \] .

Therefore, the first order Taylor series expansion of \( \Delta_{**}^3(\delta, I_1, I_2) \) is given by

\[ \Delta_{**}^3(\delta, I_1, I_2) = \left( \frac{d}{d\delta} \Delta_{**}^3(0, I_1, I_2) \right) \delta + \Delta_{**}^3(0, I_1, I_2) \]

where

\[ a_3(I_1, I_2) = \Phi(b_2(I_1+1, I_2)(I_1(I_1+1))^{1/2}) + \Phi(-b_2(I_1, I_2+1)(I_2(I_2+1))^{1/2}) \]

and

\[ b_3(I_1, I_2) = b_2(I_1+1, I_2)\Phi(b_2(I_1+1, I_2)(I_1(I_1+1))^{1/2}) \\
+ b_2(I_1, I_2+1)\Phi(-b_2(I_1, I_2+1)(I_2(I_2+1))^{1/2}) \\
+ (I_1(I_1+1))^{-1/2}\Phi(b_2(I_1+1, I_2)(I_1(I_1+1))^{1/2}) \\
- (I_2(I_2+1))^{-1/2}\Phi(-b_2(I_1, I_2+1)(I_2(I_2+1))^{1/2}) . \]

The above procedure for obtaining an approximation of \( \Delta_{**}^3(\cdot) \) can be iterated to larger values of \( n \) by substitution of the linear
approximation of $\Delta^n(\cdot)$ into the recursion equation for $\Delta^{n+1}(\cdot)$ and taking the first order Taylor series expansion about zero of the resulting quantity. Using the iterative procedure, one can obtain an approximation, $\Delta^{n+1}_n(\delta, I_1, I_2)$, of $\Delta^{n+1}(\delta, I_1, I_2)$ from the previous approximation, $\Delta^n_n(\delta, I_1, I_2)$. The above approximation of $\Delta^n(\cdot)$ will be referred to as the linear zero expansion approximation and abbreviated as LZE.

**Theorem 2.8.1:** For the LZE approximation, $a_n(I_1, I_2)\delta + b_n(I_1, I_2)$, the following properties hold:

1) $a_{n+1}(I_1, I_2) = a_n(I_1+1, I_2)\phi(-k_n(I_1+1, I_2)(I_1(I_1+1))^{1/2})$
   \[+ a_n(I_1, I_2+1)\phi(k_n(I_1, I_2+1)(I_2(I_2+1))^{1/2}), \forall n \geq 1\]

2) $b_{n+1}(I_1, I_2) = b_n(I_1+1, I_2)\phi(-k_n(I_1+1, I_2)(I_1(I_1+1))^{1/2})$
   \[+ b_n(I_1, I_2+1)\phi(k_n(I_1, I_2+1)(I_2(I_2+1))^{1/2})
   \[+ (I_1(I_1+1))^{-1/2}a_n(I_1+1, I_2)\phi(-k_n(I_1+1, I_2)(I_1(I_1+1))^{1/2})
   \[-(I_2(I_2+1))^{-1/2}a_n(I_1+1, I_2)\phi(k_n(I_1+1, I_2)(I_2(I_2+1))^{1/2}), \forall n \geq 1\]

3) $a_n(I_1, I_2) > 0; \forall n \geq 1$

where $k_n(I_1, I_2) = -\frac{b_n(I_1, I_2)}{a_n(I_1, I_2)}$ and the initial conditions are $a_1(I_1, I_2) = 1$ and $b_1(I_1, I_2) = 0$, for all $I_1, I_2$. 
Proof of i and ii:

We have already shown that the above three properties hold for n=2 in Equation 2.8.1a.

Suppose the above three properties hold for n; we need to show they hold for n+1. Let the LZE approximation for n be given by

$$\Delta^n(\delta, I_1, I_2) = a_n(I_1, I_2) \delta + b_n(I_1, I_2)$$

Since the first step in the LZE approximation is the substitution of the previous approximation into recursion Equation 2.4.12, we have that

$$\Delta^{n+1}(\delta, I_1, I_2) \equiv \int \Delta^n(\delta+z(I_1(I_1)+1))^{-1/2, I_1+1, I_2} d\phi(z) + \int \Delta^n(\delta+z(I_2(I_2)+1))^{-1/2, I_1, I_2+1} d\phi(z).$$

We denote the RHS by $\Delta^{n+1}(\delta, I_1, I_2)$.

$$\Delta^{n+1}(\delta, I_1, I_2) = \int_{S_1} a_n(I_1+1, I_2)(\delta+z(I_1(I_1+1)))^{-1/2} + b_n(I_1+1, I_2) d\phi(z) + \int_{S_2} a_n(I_1, I_2+1)(\delta+z(I_2(I_2+1)))^{-1/2} + b_n(I_1, I_2+1) d\phi(z)$$

where

$$S_1 = \{z: (\delta-k_n(I_1+1, I_2))(I_1(I_1+1))^{1/2} \geq z\}$$

and

$$S_2 = \{z: -(\delta-k_n(I_1, I_2+1))(I_2(I_2+1))^{1/2} \geq z\}.$$

Notice that the induction hypothesis that $a_n(I_1, I_2) > 0$ is used in the construction of $S_1$ and $S_2$. 
Further simplification gives that

$$\Delta_{n+1}^{a+}(\delta, I, I') = \left( a_n(I_1+1, I_2) + b_n(I_1, I_2) \right) \phi\left( (\delta - k_{n, I_1+1, I_2})(I_1(I_1+1))^{1/2} \right)$$

$$+ \left( a_n(I_1, I_2+1) + b_n(I_1, I_2+1) \right) \phi\left( (\delta - k_{n, I_1+1, I_2})(I_2(I_2+1))^{1/2} \right)$$

$$+ (I_1(I_1+1))^{-1/2} a_n(I_1, I_2) \phi\left( (\delta - k_{n, I_1+1, I_2})(I_1(I_1+1))^{1/2} \right)$$

$$- (I_2(I_2+1))^{-1/2} a_n(I_1, I_2+1) \phi\left( (\delta - k_{n, I_1+1, I_2})(I_2(I_2+1))^{1/2} \right).$$

The above equality follows from the fact that

$$\int_{-\infty}^{t} z d\phi(z) = -\phi(t).$$

We need the derivative of $\Delta_{n+1}^{a+}(\delta, I, I')$ since the second step in the LZE approximation is a Taylor series expansion of the quantity resulting from the substitution.

Now,

$$\frac{d}{d\delta} \Delta_{n+1}^{a+}(\delta, I, I') = a_n(I_1+1, I_2) \phi\left( (\delta - k_{n, I_1+1, I_2})(I_1(I_1+1))^{1/2} \right)$$

$$+ a_n(I_1, I_2+1) \phi\left( (\delta - k_{n, I_1+1, I_2})(I_2(I_2+1))^{1/2} \right)$$

$$+ (I_1(I_1+1))^{1/2} a_n(I_1, I_2) \phi\left( (\delta - k_{n, I_1+1, I_2})(I_1(I_1+1))^{1/2} \right)$$

$$- a_n(I_1+1, I_2) \phi\left( (\delta - k_{n, I_1+1, I_2})(I_1(I_1+1))^{1/2} \right)$$

$$- (I_2(I_2+1))^{1/2} a_n(I_1, I_2+1) \phi\left( (\delta - k_{n, I_1+1, I_2})(I_2(I_2+1))^{1/2} \right)$$

$$+ a_n(I_1, I_2+1) \phi\left( (\delta - k_{n, I_1+1, I_2})(I_1(I_1+1))^{1/2} \right).$$
Therefore,
\[
\frac{\partial \Delta_{n+1}^*}{\partial \delta}(\delta, I_1, I_2) = a_{n+1}(I_1+1, I_2) \phi((-k_n(I_1+1, I_2)(I_1(I_1+1)))^{1/2}) + a_{n+1}(I_1, I_2+1) \phi((-k_n(I_1, I_2+1))(I_2(I_2+1)))^{1/2})
\]

The LZE approximation of $\Delta_{n+1}^*(\cdot)$ is the first order Taylor series expansion of $\Delta_{n+1}^*(\cdot)$ and is given by
\[
\Delta_{n+1}^*(\delta, I_1, I_2) = \left[\frac{d \Delta_{n+1}^*}{d \delta^*}(0, I_1, I_2)\right] \delta + \Delta_{n+1}^*(0, I_1, I_2)
\]

where $a_{n+1}(I_1, I_2)$ and $b_{n+1}(I_1, I_2)$ are as given in i and ii.

**Proof of iii:**
From i, we have that
\[
a_{n+1}(I_1, I_2) = a_n(I_1+1, I_2) \phi((-k_n(I_1+1, I_2)(I_1(I_1+1)))^{1/2}) + a_n(I_1, I_2+1) \phi((-k_n(I_1, I_2+1))(I_2(I_2+1)))^{1/2}).
\]

But, by the induction hypothesis we know that $a_n(I_1+1, I_2)$ and $a_n(I_1, I_2+1)$ are strictly positive which implies that $a_{n+1}(I_1, I_2)$ is strictly positive.

Q.E.D.

Corollary 2.6.8 implies that we need only know $\delta^n$, the unique root of $\Delta^n(\delta, I_1, I_2)$ with respect to $\delta$, in order to characterize the optimal solution. Since the LZE approximation is a linear function of $\delta$ the root of $\Delta_{n+1}^*(\delta, I_1, I_2)$ is easily calculated and is given by
Note that the root is again unique since by part iii of Theorem 2.3.1 $a_n(I_1, I_2) > 0$. Therefore, the LZE approximation is used to define an approximation of the optimal strategy by having $\delta_n^{LZE}$ estimate $\delta^*$.

The LZE approximation can be readily calculated using the recursion equations of Theorem 2.8.1. The reparametrization of Theorem 2.4.5 implies that we need only let the difference in information needs to stage to the next be one. Therefore, the LZE approximation needs to be calculated only for integer differences in the information. Numerical computations of the LZE approximation are included in section 2.9.

The following theorem and corollary show that the LZE approximation possesses the same symmetry with respect to the information as $\Delta^n(\cdot)$. The symmetry of $\Delta^n(\cdot)$ was proven in Lemma 2.4.2.

**Theorem 2.8.2:**

i) $b_n(I_1, I_2) = -b_n(I_2, I_1)$

ii) $a_n(I_1, I_2) = a_n(I_2, I_1)$

**Proof:**

Let $n=1$; then

$b_1(I_1, I_2) = 0$, for all $I_1, I_2$

and

$a_1(I_1, I_2) = 1$, for all $I_1, I_2$.

Suppose the theorem holds for $n$; we need to show that it holds for $n+1$. 

\[
\delta_n^{LZE} = -\frac{b_n(I_1, I_2)}{a_n(I_1, I_2)}. 
\]
\[ b_{n+1}(I_1, I_2) = b_n(I_1+1, I_2) \Phi \left( \frac{b_n(I_1+1, I_2)}{a_n(I_1+1, I_2)} (I_1(I_1+1))^{1/2} \right) \]

\[ + b_n(I_1, I_2+1) \Phi \left( -\frac{b_n(I_1, I_2+1)}{a_n(I_1, I_2+1)} (I_2(I_2+1))^{1/2} \right) \]

\[ + (I_1(I_1+1))^{-1/2} a_n(I_1+1, I_2) \Phi \left( \frac{b_n(I_1+1, I_2)}{a_n(I_1+1, I_2)} (I_1(I_1+1))^{1/2} \right) \]

\[ - (I_2(I_2+1))^{-1/2} a_n(I_1, I_2+1) \Phi \left( -\frac{b_n(I_1, I_2+1)}{a_n(I_1, I_2+1)} (I_2(I_2+1))^{1/2} \right) \]

\[ = -b_n(I_2, I_1+1) \Phi \left( -\frac{b_n(I_2, I_1+1)}{a_n(I_2, I_1+1)} (I_1(I_1+1))^{1/2} \right) \]

\[ - b_n(I_2+1, I_1) \Phi \left( \frac{b_n(I_2+1, I_1)}{a_n(I_2+1, I_1)} (I_1(I_1+1))^{1/2} \right) \]

\[ + (I_1(I_1+1))^{-1/2} a_n(I_2, I_1+1) \Phi \left( -\frac{b_n(I_1, I_1+1)}{a_n(I_2, I_1+1)} (I_1(I_1+1))^{1/2} \right) \]

\[ - (I_2(I_2+1))^{-1/2} a_n(I_2, I_1, I_1) \Phi \left( \frac{b_n(I_2+1, I_1)}{a_n(I_2+1, I_1)} (I_2(I_2+1))^{1/2} \right) \]

\[ = -b_n(I_2, I_1+1) \Phi \left( -\frac{b_n(I_2, I_1+1)}{a_n(I_2, I_1+1)} (I_1(I_1+1))^{1/2} \right) \]

The first and third equalities are applications of part ii of Theorem 2.8.1 while the second follows from the induction hypothesis.

\[ a_{n+1}(I_1, I_2) = a_n(I_1+1, I_2) \Phi \left( \frac{b_n(I_1+1, I_2)}{a_n(I_1+1, I_2)} (I_1(I_1+1))^{1/2} \right) \]

\[ + a_n(I_1, I_2+1) \Phi \left( -\frac{b_n(I_1, I_2+1)}{a_n(I_1, I_2+1)} (I_2(I_2+1))^{1/2} \right) \]

\[ = a_n(I_2, I_1+1) \Phi \left( -\frac{b_n(I_2, I_1+1)}{a_n(I_2, I_1+1)} (I_1(I_1+1))^{1/2} \right) \]

\[ + a_n(I_2, I_1, I_1) \Phi \left( \frac{b_n(I_2+1, I_1)}{a_n(I_2+1, I_1)} (I_2(I_2+1))^{1/2} \right) \]

\[ = a_{n+1}(I_2, I_1) \]
The first and third inequalities are applications of part i of Theorem 2.8.1 while the second follows from the induction hypothesis.

**Corollary 2.8.3:** \( \Delta^n(\delta, I_1, I_2) = -\Delta^n(-\delta, I_2, I_1) \)

**Proof:**

\[
\Delta^n(\delta, I_1, I_2) = a_n(I_1, I_2)\delta + b_n(I_1, I_2)
= -a_n(I_2, I_1)(-\delta) - b_n(I_2, I_1)
= -\Delta^n(-\delta, I_2, I_1)
\]

Q.E.D.

The following corollary reduces the amount of calculation necessary for tabulating of the root of \( \Delta^n(\delta, I_1, I_2) \) since it is sufficient to consider the case where \( I_1 > I_2 \).

**Corollary 2.8.4:** \[ k_n(I_1, I_2) = -k_n(I_2, I_1) \]

The next theorem shows that the LZE approximation has the same limiting behavior with respect to information as \( \Delta^n(\cdot) \).

**Theorem 2.8.4:** \[
\lim_{\min(I_1, I_2) \to \infty} k_n(I_1, I_2) = 0
\]

**Proof:**

Let \( n=1 \), then \( k_1(I_1, I_2) = 0 \).
Suppose the theorem holds for $n$; we need to show that it holds for $n+1$.

$$|k_{n+1}(I_1, I_2)| \leq |k_n(I_{1}+1, I_2)| + |k_n(I_1, I_2+1)|$$

$$+ \left| \frac{(I_1(I_{1}+1))^{-1/2} \phi(-k_n(I_{1}+1, I_2)(I_1(I_{1}+1))^{1/2})}{\phi(-k_n(I_{1}+1, I_2)(I_1(I_{1}+1))^{1/2})} \right|$$

$$+ \left| \frac{(I_2(I_{2}+1))^{-1/2} \phi(k_n(I_1, I_2+1)(I_2(I_{2}+1))^{1/2})}{\phi(k_n(I_1, I_2+1)(I_2(I_{2}+1))^{1/2})} \right|$$

This inequality follows from dividing parts i and ii of Theorem 2.8.1 and shrinking the denominator by using a suitable half of the denominator.

Therefore, it would suffice to show that

$$\lim_{\min(I_1, I_2) \to \infty} \frac{(I_1(I_{1}+1))^{-1/2} \phi(-k_n(I_{1}+1, I_2)(I_1(I_{1}+1))^{1/2})}{\phi(-k_n(I_{1}+1, I_2)(I_1(I_{1}+1))^{1/2})} = 0 \quad (2.8.2)$$

and

$$\lim_{\min(I_1, I_2) \to \infty} \frac{(I_2(I_{2}+1))^{-1/2} \phi(k_n(I_1, I_2+1)(I_2(I_{2}+1))^{1/2})}{\phi(k_n(I_1, I_2+1)(I_2(I_{2}+1))^{1/2})} = 0 \quad (2.8.3)$$

We will show that Equation 2.8.2 holds by using the fact that

$$\lim_{n} x_n = x \text{ if and only if every subsequence of } <x_n> \text{ has a subsequence which converges to } x \text{ (see page 37 of Royden (1968)).}$$

Let $<x_k>$ be a sequence pair of $(I_1, I_2)$ such that $\min(I_1, I_2) \to \infty$ and let $<x_j>$ be an arbitrary subsequence of $<x_k>$. Then, $<x_j>$ must possess one
of the following two properties:

i) there exists a subsequence of \( \langle x^k \rangle \), such that
\[
k_n(I_{i+1}^2, I_1(I_{i}+1))^{1/2} \to \infty
\]

ii) there exists a number \( M \) such that \( k_n(I_{i+1}^2, I_1(I_{i}+1))^{1/2} < M \)
for all \( \langle x^k \rangle \).

Suppose that i holds. Then, for the subsequence \( \langle x^k \rangle \) we have that
\[
\lim_{\min(I_{1}, I_{2}) \to \infty} \frac{(I_{1}(I_{1}+1))^{-1/2} \phi(-k_n(I_{1}+1, I_{2})(I_{1}(I_{1}+1))^{1/2})}{\min(I_{1}, I_{2}) \to \infty, \phi(-k_n(I_{1}+1, I_{2})(I_{1}(I_{1}+1))^{1/2})} = 0
\]

The first equality results from using the tail approximation of the standard normal,
\[
\phi(t) = - \frac{1}{t} \phi(t) \text{ if } t < 0.
\]
The second equality follows from the induction hypothesis. Therefore, there exists a subsequence of \( \langle x^k \rangle \) having the desired property.

Suppose that ii holds. Then,
\[
\phi(-k_n(I_{1}+1, I_{2})(I_{1}(I_{1}+1))^{1/2}) \to \infty, \phi(-k_n(I_{1}+1, I_{2})(I_{1}(I_{1}+1))^{1/2})
\]
is bounded for all \( \langle x^k \rangle \). Therefore, for every subsequence of \( \langle x^k \rangle \).
\[ \lim_{\min(I_1, I_2) \to \infty} \frac{(I_1(I_1+1))^{-1/2} \phi(-k_n(I_1+1,I_2)(I_1(I_1+1))^{1/2})}{\Phi(-k_n(I_1+1,I_2)(I_1(I_1+1))^{1/2})} = 0. \]

Therefore, since every subsequence must possess properties i or ii and both properties imply that there exists a further subsequence such that the limit goes to zero, Equation 2.8.2 holds.

An argument similar to that given for Equation 2.8.2 shows that Equation 2.8.3 holds also.

Q.E.D.

Theorem 2.6.6 implies that \[ \lim_{\min(I_1, I_2) \to \infty} \Delta^n(0,I_1,I_2) = 0. \] Therefore, for large informations both \( \Delta^n(\cdot) \) and the LZE approximation determine the optimal strategy according to the sign of the difference of the means.

The LZE approximation uses a Taylor series expansion about a fixed point, zero, for all values of \( n \). The next two approximations will, for each value of \( n \), expand about a different point.

The next procedure for approximating \( \Delta^n(\cdot) \) will

i) substitute the linear approximation of \( \Delta^n(\cdot) \) into recursion Equation 2.4.12 and label the resulting quantity \( \Delta^{n+1}_{**}(\cdot) \),

ii) estimate the root of \( \Delta^{n+1}_{**}(\cdot) \) using one iteration of Newton's method for locating the root of an equation where the initial estimate of the root is zero,

iii) take a first order Taylor series expansion about the estimated root of \( \Delta^n_{**}(\cdot) \).
The above approximation will be referred to as the linear Newton expansion approximation and abbreviated as LNE.

**Theorem 2.8.5:** For the LNE approximation, \( a_n(I_1, I_2) \delta + b_n(I_1, I_2) \), the following properties hold:

i) \[
a_{n+1}(I_1, I_2) = a_n(I_1+1, I_2) \phi ((k^*_{n+1}(I_1, I_2) - k_n(I_1+1, I_2))(I_1(I_1+1))^{1/2})
+ a_n(I_1, I_2+1) \phi ((k^*_{n+1}(I_1, I_2) - k_n(I_1, I_2+1))(I_2(I_2+1))^{1/2})
\]

where \( k^*_{n+1}(I_1, I_2) \) is the estimated root of \( A_{n+1}^*(\cdot) \)

ii) \[
b_{n+1}(I_1, I_2) = b_n(I_1+1, I_2) \phi ((k^*_{n+1}(I_1, I_2) - k_n(I_1+1, I_2))(I_1(I_1+1))^{1/2})
+ b_n(I_1, I_2+1) \phi ((k^*_{n+1}(I_1, I_2) - k_n(I_1, I_2+1))(I_2(I_2+1))^{1/2})
+ a_n(I_1+1, I_2)(I_1(I_1+1))^{-1/2} \phi ((k^*_{n+1}(I_1+1, I_2))
- k_n(I_1+1, I_2))(I_1(I_1+1))^{1/2})
- a_n(I_1+1, I_2)(I_2(I_2+1))^{-1/2} \phi ((k^*_{n+1}(I_1+1, I_2))
- k_n(I_1+1, I_2))(I_2(I_2+1))^{1/2})
\]

iii) \[
k^*_{n+1}(I_1, I_2) = \frac{-b^*_{n+1}(I_1, I_2)}{a^*_{n+1}(I_1, I_2)}
\]
iv) \[ b_{n+1}^*(I_1, I_2) = b_n(I_1+1, I_2) \Phi(-k_n(I_1+1, I_2), ((I_1+1), 1/2)) \]

\[ + b_n(I_1, I_2+1) \Phi(k_n(I_1, I_2+1), ((I_2+1), 1/2)) \]

\[ + a_n(I_1+1, I_2) (I_1+1) ((I_1+1), 1/2)) \]

\[ - a_n(I_1, I_2+1) (I_2+1) ((I_2+1), 1/2)) \]

v) \[ a_{n+1}^*(I_1, I_2) = a_n(I_1+1, I_2) \Phi(-k_n(I_1+1, I_2), (I_1+1), 1/2)) \]

\[ + a_n(I_1, I_2+1) \Phi(k_n(I_1, I_2+1), (I_2+1), 1/2)) \]

vi) \[ a_n(I_1, I_2) > 0 \]

where \[ k_n(I_1, I_2) = \frac{b_n(I_1, I_2)}{a_n(I_1, I_2)} \] and the initial conditions are \[ a_1(I_1, I_2) = 1 \] and \[ b_1(I_1, I_2) = 0 \] for all \( I_1, I_2 \).

**Proof:** For \( n=2 \), properties i-v are readily verified using the same argument as for general \( n \), and hence, will not be included. When \( n=2 \) property vi, \( a_2(I_1, I_2) > 0 \), follows directly from property i, since \( \Phi(t) > 0 \) for all \( t \), and \( a_1(I_1, I_2) = 1 \).

Suppose the theorem holds for \( n \); we need to show that it holds for \( n+1 \).

Let the LNE approximation for \( n \) be given by

\[ \Delta_n^*(\delta, I_1, I_2) = a_n(I_1, I_2) \delta + b_n(I_1, I_2). \]

Since the first step in the LNE approximation is the substitution of the LNE approximation of \( \Delta^n(\cdot) \) into recursion Equation 2.4.12, we have that
\[ \Delta^{n+1}(\delta, I_1, I_2) \equiv \int \Delta_n^h (\delta - z (I_1 (I_1 + 1)))^{1/2} d\phi (z) + \int \Delta_n^h (\delta + z (I_2 (I_2 + 1)))^{1/2} d\phi (z). \]

The RHS of the above equation will be denoted by \( \Delta^{n+1}_{**}(\delta, I_1, I_2) \).

\[ \Delta^{n+1}_{**}(\delta, I_1, I_2) = \int (a_n (I_1 + 1, I_2) (\delta - z (I_1 (I_1 + 1)))^{1/2}) + b_n (I_1 + 1, I_2) d\phi (z) \]

Now, since \( \Delta^{n+1}_{**}(\cdot) \) for the LNE approximation has exactly the same form as \( \Delta^{n+1}_{**}(\cdot) \) for the LZE approximation we will not include the integration giving a simplification of \( \Delta^{n+1}_{**}(\cdot) \) and the calculations needed for the derivative. Therefore,

\[ \Delta^{n+1}_{**}(\delta, I_1, I_2) = (a_n (I_1 + 1, I_2) \delta + b_n (I_1 + 1, I_2)) \]

\[ \times \phi ((\delta - k_n (I_1 + 1, I_2)) (I_1 (I_1 + 1))^{1/2}) + (a_n (I_1 + 1, I_2) \delta + b_n (I_1 + 1, I_2)) \]

\[ \times \phi (-(\delta - k_n (I_1 + 1, I_2)) (I_2 (I_2 + 1))^{1/2}) + (I_1 (I_1 + 1))^{1/2} a_n (I_1 + 1, I_2) \phi ((\delta - k_n (I_1 + 1, I_2)) \]

\[ \times (I_1 (I_1 + 1))^{1/2} - (I_2 (I_2 + 1))^{1/2} a_n (I_1 + 1, I_2) \phi (-(\delta - k_n (I_1 + 1, I_2)) \]

\[ \times (I_2 (I_2 + 1))^{1/2}) \quad (2.8.4) \]
and

$$\frac{d}{d\theta} \Delta_{n+1}^{\text{**}}(\delta, I_1, I_2) = a_{n+1}(I_1, I_2) \delta ((\delta-k_n(I_1, I_2))(I_1(I_1+1))^{1/2}$$

$$+ a_{n+1}(I_1, I_2) \delta ((\delta-k_n(I_1, I_2))(I_2(I_2+1))^{1/2})$$

(2.8.5)

It should be noted that even though the form of \( \Delta_{n+1}^{\text{**}}(\cdot) \) is the same for the LNE and LZE approximations the slope and intercept are different for any \( n>2 \).

The second step in the LNE approximation is the estimation of the root of \( \Delta_{n+1}^{\text{**}}(\cdot) \) using Newton's method for locating the root of an equation where the initial estimate is zero. In general, the first iteration of Newton's method estimates the root of \( f(x) \) by

$$x_1 = x_0 - \frac{f(x_0)}{f'(x_0)}$$

where \( x_0 \) is the initial estimate. Therefore, it can be readily verified by substitution of zero into Equations 2.8.4 and 2.8.5 that the root of \( \Delta_{n+1}^{\text{**}}(\delta, I_1, I_2) \) with respect to \( \delta \) is estimated by

$$k_{n+1}^{*}(I_1, I_2) = \frac{b_{n+1}^{*}(I_1, I_2)}{a_{n+1}^{*}(I_1, I_2)}$$

where \( b_{n+1}^{*}(I_1, I_2) \) and \( a_{n+1}^{*}(I_1, I_2) \) are as given in properties iv-v.

This verifies iii-v.

In the third step of the LNE approximation we take a Taylor series expansion about the estimated root of \( \Delta_{n+1}^{\text{**}}(\delta, I_1, I_2), k_{n+1}^{*}(I_1, I_2) \).

Therefore, the LNE approximation is given by
\[
\Delta_{n+1}^{\delta}(\delta, I_1, I_2) = \left[ \frac{\partial}{\partial \delta} \Delta_{n+1}^{\delta*}(k_{n+1}^*(I_1', I_2'), I_1', I_2'] \right] (\delta - k_{n+1}^*(I_1, I_2)) \\
+ \Delta_{n+1}^{\delta*}(k_{n+1}^*(I_1', I_2'), I_1', I_2') \\
= [a_{n+1}(I_1'+1, I_2') \Phi((-k_{n+1}^*(I_1, I_2') - k_n(I_1'+1, I_2'))(I_1(I_1'+1))^{1/2}) \\
+ a_{n+1}(I_1', I_2'+1) \Phi((-k_{n+1}^*(I_1, I_2') - k_n(I_1, I_2'+1))(I_2(I_2'+1))^{1/2})]^* \\
(\delta - k_{n+1}^*(I_1, I_2')) \\
+ (a_{n+1}(I_1'+1, I_2') \Phi((-k_{n+1}^*(I_1, I_2') - k_n(I_1'+1, I_2'))(I_1(I_1'+1))^{1/2}) \\
+ a_{n+1}(I_1', I_2'+1) \Phi((-k_{n+1}^*(I_1, I_2') - k_n(I_1, I_2'+1))(I_2(I_2'+1))^{1/2})] \\
k_{n+1}^*(I_1, I_2') \\
+ b_{n}(I_1'+1, I_2') \Phi((-k_{n+1}^*(I_1, I_2') - k_n(I_1'+1, I_2'))(I_1(I_1'+1))^{1/2}) \\
+ b_{n}(I_1', I_2'+1) \Phi((-k_{n+1}^*(I_1, I_2') - k_n(I_1, I_2'+1))(I_2(I_2'+1))^{1/2}) \\
+ (I_1(I_1'+1))^{-1/2} a_{n}(I_1'+1, I_2') \Phi((-k_{n+1}^*(I_1, I_2') - k_n(I_1'+1, I_2')) \\
\times (I_1(I_1'+1))^{1/2}) \\
- (I_2(I_2'+1))^{-1/2} a_{n}(I_1, I_2'+1) \Phi((-k_{n+1}^*(I_1, I_2')) \\
- k_n(I_1, I_2')(I_2(I_2'+1))^{1/2})
\]

using Equations 2.8.4 and 2.8.5. The above equations simplify to the following form

\[
\Delta_{n+1}^{\delta}(\delta, I_1, I_2) = a_{n+1}(I_1, I_2) \delta + b_n(I_1+I_2)
\]

where \(a_{n+1}(I_1, I_2)\) and \(b_{n+1}(I_1, I_2)\) are as given in properties i and ii.
Now, we will verify property vi that $a_{n+1}(I_1, I_2) > 0$. From property i we have

$$a_{n+1}(I_1, I_2) = a_n(I_{n+1}, I_2)\phi((k^*_n(I_1, I_2) - k_n(I_1, I_2))(I_1(I_1 + 1))^{1/2})$$

$$+ a_n(I_1, I_2 + 1)\phi(-(k^*_n(I_1, I_2) - k_n(I_1, I_2 + 1))(I_2(I_2 + 1))^{1/2}).$$

But, by the induction hypothesis we have that $a_n(\cdot)$ is strictly positive. Therefore, $a_{n+1}(\cdot) > 0$.

Q.E.D.

For reasons similar to those given for the L2E approximation, an approximation of the optimal strategy is given by substituting the root of $a^{n}(\cdot)$, $\hat{\delta}^{LNE}$, for the root of $a^{n}(\cdot)$, $\hat{\delta}^*$, in Corollary 2.6.8 where

$$\hat{\delta}^{LNE} = -\frac{b_n(I_1, I_2)}{a_n(I_1, I_2)}.$$

Numerical computations of the LNE approximation are given in section 2.9.

The following theorem and corollary show that the LNE approximation proposes the same symmetry with respect to the informations as $a_n(\cdot)$. Statements i-iii of Theorem 2.8.6 are included for clarity, since they are needed in the proofs of iv and v.

**Theorem 2.8.6:**

i) $a_n^*(I_1, I_2) = a_n^*(I_2, I_1)$

ii) $b_n^*(I_1, I_2) = -b_n^*(I_2, I_1)$

iii) $k_n^*(I_1, I_2) = -k_n^*(I_2, I_1)$

iv) $a_n(I_1, I_2) = a_n(I_2, I_1)$

v) $b_n(I_1, I_2) = -b_n(I_2, I_1)$
Proof: Note that if $n=2$,

$$a_2^L(I_1, I_2) = \frac{d}{d\xi} \Delta^2(0, I_1, I_2)$$

$$= a_2^{LZE}(I_1, I_2)$$

and

$$b_2^L(I_1, I_2) = \Delta^2(0, I_1, I_2)$$

$$= b_2^{LZE}(I_1, I_2)$$

where $a_2^{LZE}(\cdot)$ and $b_2^{LZE}(\cdot)$ are the slope and intercept of the LZE approximation, respectively. Therefore, $a_2^L(\cdot)$ and $b_2^L(\cdot)$ have the desired symmetry with respect to the information since $a_2^{LZE}(\cdot)$ and $b_2^{LZE}(\cdot)$ have it by Theorem 2.8.2. Thus,

$$k_2^L(I_1, I_2) = \frac{-b_2(I_1, I_2)}{a_2(I_1, I_2)}$$

$$= \frac{b_2(I_2, I_1)}{a_2(I_2, I_1)}$$

$$= -k_2^L(I_2, I_1)$$

(2.8.6)

This verifies properties i-iii for $n=2$.

Now, we will verify properties iv and v for $n=2$.

$$a_2(I_1, I_2) = \phi(k_2^L(I_1, I_2)(I_1(I_1+1))^{1/2}) + \phi(-k_2^L(I_1, I_2)(I_2(I_2+1))^{1/2})$$

$$= \phi(-k_2^L(I_2, I_1)(I_1(I_1+1))^{1/2}) + \phi(k_2^L(I_2, I_1)(I_2(I_2+1))^{1/2})$$

$$= a_2(I_2, I_1)$$

The first and third equalities follow from part i of Theorem 2.8.5 while the second equality is implied by Equation 2.8.6.
\[ b_2(I_1, I_2) = (I_1(I_1+1))^{-1/2} \phi(k_2^*(I_1, I_2)(I_1(I_1+1)))^{1/2} \]
\[- (I_2(I_2+1))^{-1/2} \phi(-k_2^*(I_1, I_2)(I_2(I_2+1)))^{1/2} \]
\[= - (I_2(I_2+1))^{-1/2} \phi(k_2^*(I_2, I_1)(I_2(I_2+1)))^{1/2} \]
\[- (I_1(I_1+1))^{-1/2} \phi(-k_2^*(I_2, I_1)(I_1(I_1+1)))^{1/2} \]
\[= - b_2(I_2, I_1) \]

The first and third equalities follow from part ii of Theorem 2.8.5 while the second equality is implied by Equation 2.8.6.

This verifies properties iv and v for \( n=2 \).

Suppose the theorem holds for \( n \) we need to show that it holds for \( n+1 \).

We will first verify properties i-iii for \( n+1 \).

\[ a^*_{n+1}(I_1, I_2) = a_n(I_1+1, I_2) \phi(k_n(I_1+1, I_2)(I_1(I_1+1)))^{1/2} \]
\[+ a_n(I_1, I_2+1) \phi(-k_n(I_1, I_2+1)(I_2(I_2+1)))^{1/2} \]
\[= a_n(I_2, I_1+1) \phi(-k_n(I_2, I_1+1)(I_1(I_1+1)))^{1/2} \]
\[+ a_n(I_2+1, I_1) \phi(k_n(I_2+1, I_1)(I_2(I_2+1)))^{1/2} \]
\[= a^*_{n+1}(I_2, I_1) \]

and

(2.8.7)
\[ b_{n+1}(I_1, I_2) = b_n(I_1+1, I_2) \phi(k_n(I_1+1, I_2)(I_1(I_1+1))^{1/2}) \]

\[ + b_n(I_1, I_2+1) \phi(-k_n(I_1, I_2+1)(I_2(I_2+1)))^{1/2} \]

\[ + a_n(I_1+1, I_2)(I_1(I_1+1))^{-1/2} \phi(k_n(I_1+1, I_2)(I_1(I_1+1)))^{1/2} \]

\[ - a_n(I_1, I_2+1)(I_2(I_2+1))^{-1/2} \phi(-k_n(I_1, I_2+1)(I_2(I_2+1)))^{1/2} \]

\[ = -b_n(I_2, I_1+1) \phi(-k_n(I_2, I_1+1)(I_1(I_1+1)))^{1/2} \]

\[ - b_n(I_2+1, I_1) \phi(k_n(I_2+1, I_1)(I_2(I_2+1)))^{1/2} \]

\[ + a_n(I_2, I_1+1)(I_1(I_1+1))^{-1/2} \phi(-k_n(I_2, I_1+1)(I_1(I_1+1)))^{1/2} \]

\[ - a_n(I_2+1, I_1)(I_2(I_2+1))^{-1/2} \phi(k_n(I_2+1, I_1)(I_2(I_2+1)))^{1/2} \]

\[ = -b_{n+1}(I_2, I_1) \quad (2.8.9) \]

Equations 2.8.7 and 2.8.8 follow from parts iv and v of Theorem 2.8.5 and the induction hypothesis.

\[ k_{n+1}(I_1, I_2) = -\frac{b_{n+1}(I_1, I_2)}{a_{n+1}(I_1, I_2)} \]

\[ = \frac{b_{n+1}(I_2, I_1)}{a_{n+1}(I_2, I_1)} \quad (2.8.9a) \]

Equation 2.8.9a follows from Equations 2.8.7 and 2.8.8. This verifies properties i–iii for n+1. Now, we will verify iv and v using iii.
The first and third equalities follow from part i of Theorem 2.8.5 while the second is an application of the induction hypothesis.

Since the proof that

$$b_{n+1}(I_1, I_2) = -b_{n+1}(I_2, I_1)$$

(2.8.10)

is similar to that given by Equation 2.8.9, we will not prove Equation 2.8.10.

Q.E.D.

The following corollary shows that the LNE approximation has the same symmetry with respect to information as $\Delta^n(\cdot)$. 

Corollary 2.8.7: $\Delta^n(\delta, I_1, I_2) = -\Delta^n(-\delta, I_1, I_2)$ 

where $\Delta^n(\cdot)$ is the LNE approximation.

**Proof:**

$$\Delta^n(\delta, I_1, I_2) = a_n(I_1, I_2) \delta + b_n(I_1, I_2)$$

$$= -a_n(I_2, I_1) (-\delta) + b_n(I_2, I_1)$$

$$= -\Delta^n(-\delta, I_1, I_2).$$
The following corollary, which is similar to Corollary 2.8.4, reduces the amount of calculation necessary for tabulating the LNE approximation.

**Corollary 2.8.8:** \( k_n(I_1, I_2) = -k_n(I_2, I_1) \)

where

\[
k_n(I_1, I_2) = -\frac{b_n(I_1, I_2)}{a_n(I_1, I_2)}.
\]

The next theorem shows that for large informations, the strategies associated with \( \Delta^n(.) \) and the LNE approximation determine which source to sample according to the sign of the difference of the means. The proof of Theorem 2.8.9 is the same as the proof of 2.8.4 if we substitute

\[
(k^*_{n+1}(I_1, I_2) - k_n(I_1+1, I_2)) (I_1(I_1+1))^{1/2}
\]

for

\[
-k_n(I_1+1, I_2) (I_1(I_1+1))^{1/2}.
\]

Hence, no proof will be given.

**Theorem 2.8.9:** \( \lim_{\min(I_1, I_2) \to \infty} k_n(I_1, I_2) = 0 \)

where

\[
k_n(I_1, I_2) = -\frac{b_n(I_1, I_2)}{a_n(I_1, I_2)}.
\]
The LNE approximation uses one iteration of Newton's method for estimating the root of a function to provide a point about which to expand. The next approximation will expand about $k_n(I_1', I_2)$, the root of the previous approximation using the current informations.

The next approximation for $\Delta^i_n(\cdot)$ can be described as follows:

i) substitute the linear approximation of $\Delta^i_n(\cdot)$ into the recursion Equation 2.4.12 and label the resulting quantity $\Delta^{i+1}_n(\cdot)$

ii) take a first order Taylor series expansion about $k_n(I_1', I_2)$

where $k_n(I_1', I_2) = \frac{b_n(I_1', I_2)}{a_n(I_1', I_2)}$.

The above approximation will be referred to as the linear root expansion approximation and abbreviated as LRE.

Theorem 2.8.9: For the LRE approximation, $a_n(I_1', I_2)\delta + b_n(I_1', I_2)$, the following properties hold:

i) $a_{n+1}(I_1', I_2) = a_n(I_1'+1, I_2)\phi((k_n(I_1', I_2)-k_n(I_1'+1, I_2))(I_1(I_1'+1))^{1/2})$

$+ a_n(I_1', I_2+1)\phi((-k_n(I_1', I_2)-k_n(I_1', I_2+1))(I_2(I_2+1))^{1/2})$

ii) $b_{n+1}(I_1', I_2) = b_n(I_1'+1, I_2)\phi((k_n(I_1', I_2)-k_n(I_1'+1, I_2))(I_1(I_1'+1))^{1/2})$

$+ b_n(I_1', I_2+1)\phi((-k_n(I_1', I_2)-k_n(I_1', I_2+1))(I_2(I_2+1))^{1/2})$

$+ (I_1(I_1'+1))^{-1/2}a_n(I_1'+1, I_2)\phi((k_n(I_1', I_2)-k_n(I_1'+1, I_2))(I_1(I_1'+1))^{1/2})$

$- (I_2(I_2'+1))^{1/2}a_n(I_1', I_2+1)\phi((-k_n(I_1', I_2)-k_n(I_1', I_2+1))(I_1(I_1'+1))^{1/2})$

iii) $a_n(I_1', I_2) > 0$
where
\[ k_n(I_1, I_2) = \frac{b_n(I_1, I_2)}{a_n(I_1, I_2)} \]

and the initial conditions are \( a_1(I_1, I_2) = 1 \) and \( b_1(I_1, I_2) = 0 \).

Proof:

For \( n = 2 \) properties i and ii are readily verified using the same argument as for general \( n \) and hence, will not be included. When \( n = 2 \), property iii follows directly from property i, since \( I(t) > 0 \) for all \( t \) and \( a_1(I_1, I_2) = 1 \).

Let the LRE approximation for \( n \) be given by

\[ \Delta^R_n(I_1, I_2) = a_n(I_1, I_2)\delta + b_n(I_1, I_2). \]

Since the first step in the LRE approximation is the substitution of the previous approximation, \( \Delta^R_n(\cdot) \), into recursion Equation 2.4.12.

\[ \Delta^{n+1}(\delta, I_1, I_2) = \left[ \Delta^R_n(\delta - z(I_1(I_1+1))^{-1/2}, I_1 + 1, I_2) \right] + \int \Delta^R_n(\delta + z(I_2(I_2+1))^{-1/2}, I_1, I_2 + 1) \, d\delta(z) \]

The RHS of the above equation will be denoted by \( \Delta^+_{n+1}(\delta, I_1, I_2) \).

\[ \Delta^+_{n+1}(\delta, I_1, I_2) = \int \left[ a_n(I_1+1, I_2)(\delta - z(I_1(I_1+1)))^{-1/2} \right] + b_n(I_1+1, I_2) \, d\delta(z) \]

\[ + \int (a_n(I_1, I_2+1)(\delta + z(I_2(I_2+1)))^{-1/2} + b_n(I_1, I_2+1)) \, d\delta(z) \]
Since $\Delta_{n+1}^n(\cdot)$ for the LRE approximation has exactly the same form as $\Delta_{n+1}^n(\cdot)$ for the LZE approximation, we will not include the integration giving a simplification of $\Delta_{n+1}^n(\cdot)$ and the calculations needed for the derivative. The second step in the LRE approximation is a first order Taylor series expansion about $k_n(I_1, I_2)$. Therefore, the LRE approximation is given by

\[
\Delta_{n+1}^n(\delta, I_1, I_2) = \frac{d}{d\delta} \Delta_{n+1}^n(k_n(I_1, I_2), I_1, I_2)[\delta - k_n(I_1, I_2)] \]

\[+ \Delta_{n+1}^n(k_n(I_1, I_2), I_1, I_2). \tag{2.8.11} \]

Equation 2.8.11 simplifies to the following form

\[
\Delta_{n+1}^n(\delta, I_1, I_2) = a_{n+1}(I_1, I_2)\delta + b_{n+1}(I_1, I_2)
\]

where $a_{n+1}(I_1, I_2)$ and $b_{n+1}(I_1, I_2)$ are as given in i and ii.

Property iii follows directly from property i and the induction hypothesis that $a_n(I_1, I_2) > 0$.

Q.E.D.

Since the LNE and LRE approximations have similar structures, the following two theorems and related corollaries will be stated without proof. The LNE approximation recursion equations for $a_{n+1}(I_1, I_2)$ and $b_{n+1}(I_1, I_2)$ are the same as those for the LRE approximation if we substitute $k_n(I_1, I_2)$ for $k_{n+1}^*(I_1, I_2)$.

Theorem 2.8.10 shows that the LRE approximation has the same symmetry with respect to the informations as $\Delta^n(\cdot)$. The proof of this theorem is the same as that of Theorem 2.8.6, if we substitute
\( k_n(I_1, I_2) \) for \( k_{n+1}(I_1, I_2) \), beginning at Equation 2.8.9 and noting that

\[ k_n(I_1, I_2) = -k_n(I_2, I_1). \]

**Theorem 2.8.10:** For the LRE approximation,

i) \( a_n(I_1, I_2) = a_n(I_2, I_1) \)

ii) \( b_n(I_1, I_2) = -b_n(I_2, I_1) \).

**Corollary 2.8.11:** \( \Delta^n(\delta, I_1, I_2) = -\Delta^n(-\delta, I_2, I_1) \) where \( \Delta^n(\cdot) \) is the LRE approximation.

**Corollary 2.8.12:** For the LRE approximation,

\[ k_n(I_1, I_2) = -k_n(I_2, I_1). \]

**Theorem 2.8.13** shows that the LRE approximation has the same limiting behavior with respect to information as \( \Delta^n(\cdot) \). The proof of this theorem is the same as that for Theorem 2.8.4 if we substitute

\[ (k_n(I_1, I_2) - k_n(I_1, I_2 + 1))(I_1(I_1 + 1))^{1/2} \]

for

\[ -k_n(I_1, I_2)(I_2(I_2 + 1))^{1/2}. \]

**Theorem 2.8.13:** For the LRE approximation,

\[ \lim_{\min(I_1, I_2) \to \infty} k_n(I_1, I_2) = 0 \]

Therefore, for large information the strategies associated with \( \Delta^n(\cdot) \) and the LRE approximation determine which source to sample according to the sign of the difference of the means.
2.9. Numerical Computations

This section is devoted to the numerical comparison of the three linear approximations of section 2.8, the bound $h_n(I_2, I_1)$ of Corollary 2.7.6 and the source differential function $\Delta^n(\delta, I_1, I_2)$ itself. All comparisons are made in terms of the roots of these expressions. The reason for this criterion is that Corollary 2.6.8 implies that the strategy is determined by the root.

The comparisons of the roots are done for integer values of the informations. Integer values of the informations are suggested by Corollary 2.4.6 which states that we need only consider integer differences in the informations. Lemma 2.6.9 implies that the root of $\Delta^n(\cdot)$ is symmetric with respect to the informations while Corollaries 2.8.2, 2.8.6 and 2.8.8 imply that the approximations possess the same symmetry. Therefore, only the case where $I_1 > I_2$ is considered.

Equation 2.5.1 provides a closed formula for calculating $\Delta^2(\cdot)$ while $\Delta^3(\cdot)$ is computed using recursion Equation 2.4.12, and the closed form of $\Delta^2(\cdot)$. The largest value of $n$ for which computations are readily performed is $n=3$, since each unit increase in $n$ requires an additional level of integration.

The tabulated roots of $\Delta^n(\cdot)$ in Table 2.9.1 possess three decimal place accuracy and are such that $\Delta^n(\cdot) < 10^{-5}$. Since all the approximations have the structure $a_n(I_1, I_2) + b_n(I_1, I_2)$, the root $k_n(\cdot)$ is given by $-\frac{b_n(I_1, I_2)}{a_n(I_1, I_2)}$. 


Theorem 2.6.6 implies that $\Delta^n(0, I_1, I_2) \to 0$ as $\min(I_1, I_2) \to \infty$. Table 2.9.1 illustrates that the convergence of Theorem 2.6.6 is quite rapid for small values of $n$. Also, the unproven proposition that for fixed information the root is an increasing function of $n$ holds for $n=2, 3$, if $I_1 > I_2$. This proposition is plausible because as $n$ increases, the repercussions of not locating the superior source also increases.

Notice that the bound, the LZE approximation, and the LRE approximations all provide the same estimate of the root for $n=2$. The bound and the LZE approximation are the same since $\delta - \Delta^n(\delta, I_1, I_2)$ is maximized at $\delta=0$ and the LZE approximation is a Taylor series expansion about zero where the slope is one. The reason that the LZE and LRE approximations provide the same estimate is that the initial conditions imply $k_1(I_1, I_2) = 0$.

For small values of $n$, the LNE approximation appears to be superior to both the LNE and LZE approximations. The LNE approximation possesses more accuracy than the other two approximations for all values of information examined. The LZE approximation is unable to react to changing amounts of information since it expands about the same point, zero, for any information and in fact violates the upper bound in some instances. The LRE approximation uses different expansion points for each set of informations, but the points do not appear to move far enough away from zero. All the approximations eventually overestimate. This may be due to the fact that the slope $a_n(I_1, I_2)$ eventually differ too much from the asymptotic slope of one. This conjecture is supported
Table 2.9.1. Roots of $A^n(\cdot)$, of the linear approximations LZE, LNE, and LRE and of the upper bound function $h^n(\cdot)$

<table>
<thead>
<tr>
<th>$I_1$</th>
<th>$I_2$</th>
<th>$n$</th>
<th>$\Delta^n(\cdot)$</th>
<th>LZE</th>
<th>LNE</th>
<th>LRE</th>
<th>Upper bound</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>1</td>
<td>2</td>
<td>.116</td>
<td>.119</td>
<td>.116</td>
<td>.119</td>
<td>.119</td>
</tr>
<tr>
<td></td>
<td></td>
<td>3</td>
<td>.177</td>
<td>.225</td>
<td>.185</td>
<td>.217</td>
<td>.286</td>
</tr>
<tr>
<td>3</td>
<td>1</td>
<td>2</td>
<td>.157</td>
<td>.167</td>
<td>.157</td>
<td>.167</td>
<td>.167</td>
</tr>
<tr>
<td></td>
<td></td>
<td>3</td>
<td>.243</td>
<td>.338</td>
<td>.258</td>
<td>.305</td>
<td>.360</td>
</tr>
<tr>
<td>5</td>
<td>1</td>
<td>2</td>
<td>.184</td>
<td>.209</td>
<td>.184</td>
<td>.209</td>
<td>.209</td>
</tr>
<tr>
<td></td>
<td></td>
<td>3</td>
<td>.293</td>
<td>.481</td>
<td>.321</td>
<td>.384</td>
<td>.430</td>
</tr>
<tr>
<td>3</td>
<td>2</td>
<td>2</td>
<td>.047</td>
<td>.048</td>
<td>.047</td>
<td>.048</td>
<td>.048</td>
</tr>
<tr>
<td></td>
<td></td>
<td>3</td>
<td>.073</td>
<td>.090</td>
<td>.080</td>
<td>.089</td>
<td>.121</td>
</tr>
<tr>
<td>5</td>
<td>2</td>
<td>2</td>
<td>.086</td>
<td>.090</td>
<td>.086</td>
<td>.090</td>
<td>.090</td>
</tr>
<tr>
<td></td>
<td></td>
<td>3</td>
<td>.136</td>
<td>.183</td>
<td>.149</td>
<td>.168</td>
<td>.191</td>
</tr>
<tr>
<td>7</td>
<td>2</td>
<td>2</td>
<td>.100</td>
<td>.110</td>
<td>.100</td>
<td>.110</td>
<td>.110</td>
</tr>
<tr>
<td></td>
<td></td>
<td>3</td>
<td>.161</td>
<td>.241</td>
<td>.179</td>
<td>.206</td>
<td>.225</td>
</tr>
<tr>
<td>7</td>
<td>5</td>
<td>2</td>
<td>.019</td>
<td>.020</td>
<td>.019</td>
<td>.020</td>
<td>.020</td>
</tr>
<tr>
<td></td>
<td></td>
<td>3</td>
<td>.031</td>
<td>.038</td>
<td>.035</td>
<td>.037</td>
<td>.045</td>
</tr>
<tr>
<td>9</td>
<td>5</td>
<td>2</td>
<td>.030</td>
<td>.031</td>
<td>.030</td>
<td>.031</td>
<td>.031</td>
</tr>
<tr>
<td></td>
<td></td>
<td>3</td>
<td>.049</td>
<td>.061</td>
<td>.055</td>
<td>.059</td>
<td>.066</td>
</tr>
</tbody>
</table>

by the fact that the bound $h^n(\cdot)$, which uses an iterated maximum with fixed slope one, is reasonably tight.
3. SINGULAR NORMAL PRIOR

3.1. Background

In this chapter, we will investigate the problem of how to sequentially decide which of two dependent normal sources to sample when the objective is to minimize the expected sum of the observations. The sources are dependent since the mean vector is restricted to lie on a line. The situation considered here is similar to that of Feldman (1962).

The second section reviews the normal posterior and marginal distributions when a singular prior is assumed. The next section establishes concepts analogous to those of sections 2.3 and 2.4 while the last section develops the optimal strategy for the special case where $J_1 = J_2 = J$.

3.2. Bayesian Distributions for the Singular Prior Case

Consider the two information sources $X$ and $Y$ for the two-armed bandit singular prior problem of Chapter 1. Let $x$ denote an observation from source $X$ and $y$ an observation from source $Y$.

Assume that

i) $x \sim \mathcal{N}(0, \frac{1}{J_1})$; $J_1 > 0$

ii) $y \sim \mathcal{N}(-\theta, \frac{1}{J_2})$; $J_2 > 0$ \hspace{1cm} (3.2.1)

iii) $J_1, J_2$ are known.
The Bayesian approach calls for a prior distribution on the unknown parameter \( \theta \). A reasonable prior distribution is a normal distribution. Thus, the prior distribution on \( \theta \) satisfies

\[ \begin{align*}
&i) \quad \theta \sim N(\mu, \frac{1}{I}), \quad I > 0 \\
&ii) \quad \mu \text{ and } \frac{1}{I} \text{ known.}
\end{align*} \] (3.2.2)

We parametrize in terms of the information for the same reasons as given in Chapter 2.

A model that has the same conditional, marginal, and posterior distributions as the situation given in Equations 3.2.1 and 3.2.2 is

\[
\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 1 \\ -1 \end{pmatrix} \theta + \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \end{pmatrix}
\]

where \( \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \end{pmatrix} \sim N\left( \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} J_1^{-1} & 0 \\ 0 & J_2 \end{pmatrix} \right) \) (3.2.3)

and

\( \theta \sim N(\mu, I^{-1}) \).

The above model has the advantage that the posterior and marginal distributions are much easier to calculate.

Theorem 3.2.1:

For the model given in Equation 3.2.3,

\[
\begin{align*}
&i) \quad \theta | x \sim N\left( E(\theta | x), \begin{pmatrix} V(\theta | x) - V(\theta | x) & -V(\theta | x) J_2^{-1} V(\theta | x) \\ -V(\theta | x) J_2^{-1} V(\theta | x) & V(\theta | x) \end{pmatrix} \right)
\end{align*}
\]
where

\[ E(\theta|x) = \frac{I}{I+J_1} \mu + \frac{J_1}{I+J_1} x \]

and

\[ V(\theta|x) = (I+J_1)^{-1}. \]

\[ \theta|x \sim N((E(\theta|y), V(\theta|y), V(\theta|y)) \quad \text{where} \]

\[ E(\theta|y) = \frac{I}{I+J_2} \mu - \frac{J_2}{I+J_2} y, \]

and

\[ V(\theta|y) = (I+J_2)^{-1}. \]

\[ \mu \sim N(-\mu, I^{-1}+J_1^{-1}) \]

\[ \nu \sim N(-\mu, I^{-1}+J_2^{-1}) \]

**Proof of i:**

Equation 3.2.3 implies that

\[
\begin{pmatrix}
\theta \\
y \\
x
\end{pmatrix} \sim N
\begin{pmatrix}
\mu \\
-\mu \\
\mu
\end{pmatrix},
\begin{pmatrix}
I^{-1} & -I^{-1} & I^{-1} \\
-I^{-1} & I^{-1}+J_2^{-1} & -I^{-1} \\
-I^{-1} & -I^{-1} & I^{-1}+J_1^{-1}
\end{pmatrix}.
\]

But, by standard results for the normal family we have that
where

\[ \mu = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix} + \begin{pmatrix} I^{-1} \\ -I^{-1} \end{pmatrix} \begin{pmatrix} I^{-1} + J_1^{-1} \end{pmatrix}^{-1} (x - \mu) \]

\[ = \left[ \mu + \left( I^{-1} \right) \begin{pmatrix} I^{-1} + J_1^{-1} \end{pmatrix}^{-1} (x - \mu) \right] (0 \ 0)_1 \]

\[ = \begin{pmatrix} I \\ I + J_1 \end{pmatrix} \mu + \begin{pmatrix} J_1 \\ I + J_1 \end{pmatrix} x (0 \ 0)_1 \]

\[ = \begin{pmatrix} E(\theta | x) \\ -E(\theta | x) \end{pmatrix} \]

and

\[ \nu = \begin{pmatrix} I^{-1} \\ -I^{-1} \end{pmatrix} - \begin{pmatrix} I^{-1} \end{pmatrix} \begin{pmatrix} I^{-1} + J_1^{-1} \end{pmatrix}^{-1} (I^{-1}, -I^{-1}) \]

\[ = (I^{-1} - (I^{-1} + J_1^{-1})) (0 \ 0) \begin{pmatrix} 1 \ 1 \end{pmatrix} \]

\[ = (I + J_1^{-1}) (0 \ 0) \begin{pmatrix} 1 \ 1 \end{pmatrix} \]

\[ = \begin{pmatrix} V(\theta | x) \\ -V(\theta | x) \end{pmatrix} \begin{pmatrix} J_2^{-1} + V(\theta | x) \end{pmatrix} \]

Q.E.D.

The proof of ii is similar to that for i and therefore, will not be given. Properties iii and iv are direct consequences of Equation 3.2.4.

Q.E.D.
Simple extensions of Theorem 3.2.1 show that the posterior distribution of $\theta$ is $N(E(\theta|x,y), V(\theta|x,y))$ after observations $x_1, x_2, \ldots, x_n$ and $y_1, y_2, \ldots, y_n$ have been taken where

$$E(\theta|x,y) = (I + n_1 J_1 + n_2 J_2)^{-1} (I\mu + n_1 J_1 \bar{x} - n_2 J_2 \bar{y})$$

and

$$V(\theta|x,y) = (I + n_1 J_1 + n_2 J_2)^{-1}.$$

### 3.3. The Singular Prior Source

#### Differential Function

The development for the singular prior case is similar to that for the independent prior case and hence, concepts analogous to those of sections 2.3 and 2.4 will be used without full justification. The definition of action space and strategy are the same as those given in section 2.3. We will also assume the same loss function which implies that the definitions of risk, Bayes risk, and optimal strategy can remain unchanged.

Let

$$V^n(\mu, I) = \min_{s \in S^n} R(\xi, s)$$

(3.3.1)

where $S^n$ is the set of strategies taking a total of $n$ observations and $\xi(\cdot)$ is the prior distribution of $\theta$.

As pointed out in section 3.2, the prior distribution of $\theta$ here is assumed to be $N(\mu, I^{-1})$. Notice that the Bayes risk of Equation 3.3.1 is a function of the prior parameters $\mu$ and $I$. 
Let
\[ V^n_x(\mu, I) = \min_{s_n \in S^n_x} R(\xi, s_n) \tag{3.3.2} \]
where
\[ S^n_x = \{s_n : s_1 = 1\}. \]
and
\[ V^n_y(\mu, I) = \min_{s_n \in S^n_y} R(\xi, s_n) \tag{3.3.3} \]
where
\[ S^n_y = \{s_n : s_1 = 0\}. \]

\( V^n_x(\cdot) \) is simply the Bayes risk given that we are restricted to sampling source \( x \) first; \( V^n_y(\cdot) \) has a similar interpretation.

If we substitute the posterior and marginal distributions of section 3.2 for those of section 2.2, the following recursion equations hold for the same reasons as given in section 2.3.
\[ V^n_x(\mu, I) = \mu + \int V^n \left( \frac{1}{I+J_1} \mu + \frac{1}{I+J_1} x, I+J_1 \right) dP(x) \tag{3.3.4} \]
and
\[ V^n_y(\mu, I) = -\mu + \int V^n \left( \frac{I}{I+J_2} \mu - \frac{J_2}{I+J_2} y, I+J_2 \right) dP(y) \tag{3.3.5} \]
where \( P(x) \) and \( P(y) \) are the distribution functions of \( N(\mu, I^{-1} + J_1^{-1}) \) and \( N(\mu, I^{-1} + J_2^{-1}) \) random variables, respectively.

The optimal strategy given in the following theorem is the analog of the strategy found in Theorem 2.3.2.
Theorem 3.3.1: An optimal strategy is given by

i) sample source X if $\Delta^n(\mu^*, I^*) > 0$

ii) sample source Y if $\Delta^n(\mu^*, I^*) < 0$

iii) sample either source if $\Delta^n(\mu^*, I^*) = 0$

where

$\Delta^n(\mu, I) = V^n_y(\mu, I) - V^n_x(\mu, I)$

$n = \text{number of observations still to be observed}$

$\mu^* = \text{current posterior mean}$

$\mu^* = (I+n_1 J_1 + n_2 J_2)^{-1}(I + n_1 J_1 \bar{x} + n_2 J_2 \bar{y})$

$n_1 = \text{number of observations taken on X}$

$n_2 = \text{number of observations taken on Y}$

$\bar{x} = \text{mean of the observations taken on X}$

$\bar{y} = \text{mean of the observations taken on Y}$

$I^* = I + n_1 J_1 + n_2 J_2$.

Theorem 3.3.1 suggests that $\Delta^n(\mu, I)$, the source differential function, is the quantity that should receive further attention. A functional relationship for $\Delta^n(\cdot)$ similar to 3.3.4 and 3.3.5 can be established as follows:

**Theorem 3.3.2:**

$\Delta^{n+1}(\mu, I) = \left\{ \Delta^n\left(\frac{I}{I+J_1} \mu + \frac{J_1}{I+J_1} x, I+J_1\right)^+ dP(x) + \Delta^n\left(\frac{I}{I+J_2} \mu - \frac{J_2}{I+J_2} y, I+J_2\right)^- dP(y) \right\}$
where \( P(x) \) and \( P(y) \) are the distribution functions of \( N(\mu, I^{-1} + J^{-1}_1) \) and \( N(\mu, I^{-1} + J^{-1}_2) \) random variables, respectively.

Proof:

Define

\[
V^n_{xy}(\mu, I) = \min_{s \in S^n_{xy}} R(\xi, s_{-n})
\]

where

\[
S^n_{xy} = \{s_{-n} : s_1 = 1 \text{ and } s_2 = 0\}
\]

and

\[
V^n_{yx}(\mu, I) = \min_{s \in S^n_{yx}} R(\xi, s_{-n})
\]

where

\[
S^n_{yx} = \{s_{-n} : s_1 = 0 \text{ and } s_2 = 1\}
\]

Notice that

\[
V^n_{xy}(\mu, I) = V^n_{yx}(\mu, I), \tag{3.3.6}
\]

and

\[
V^{n+1}_{xy}(\mu, I) = \mu + \int_{y} V^n_{y(I+I)^{-1}_1} \mu + \frac{I}{I+J_2} x, I+J_2) dP(x). \tag{3.3.7}
\]

and

\[
V^{n+1}_{yx}(\mu, I) = -\mu + \int_{x} V^n_{x(I+J)^{-1}_2} \mu - \frac{I}{I+J_1} y, I+J_2) dP(y) \tag{3.3.8}
\]

Now, if we combine 3.3.7 and 3.3.8, we have the following series of equalities.
\[ \Delta^{n+1}(\mu, I) = v^n_y(\mu, I) - v^n_{yx}(\mu, I) + v^n_{xy}(\mu, I) - v^n_x(\mu, I) \]
\[ = \int v^n\left(\frac{I}{I+J_2} - \frac{J_2}{I+J_2}, y, I+J_2\right) \]
\[ - v^n\left(\frac{I}{I+J_2} - \frac{J_2}{I+J_2}, y, I+J_2\right) dP(y) \]
\[ + \int v^n\left(\frac{I}{I+J_2}, y, I+J_2\right) \]
\[ - v^n\left(\frac{I}{I+J_1}, y, I+J_1\right) dP(x) \]
\[ = \int \Delta^n\left(\frac{I}{I+J_2} - \frac{J_2}{I+J_2}, y, I+J_2\right) dP(y) \]
\[ + \int \Delta^n\left(\frac{I}{I+J_1} + \frac{J_2}{I+J_2}, x, I+J_2\right) dP(x) \]

Q.E.D.

If we standardize the conclusion of Theorem 3.3.2 we have that
\[ \Delta^{n+1}(\mu, I) = \int \Delta^n(\mu+\frac{J_1}{I(I+J_1)}\frac{1}{\sqrt{2}}, I+J_1) d\Phi(z) \]
\[ + \int \Delta^n(\mu-\frac{J_2}{I(I+J_2)}\frac{1}{\sqrt{2}}, I+J_2) d\Phi(z). \quad (3.3.9) \]

The following theorem shows that \( \Delta^n(\cdot) \) is a strictly decreasing function of \( \mu \). For the independent prior case, \( \Delta^n(\cdot) \) is an increasing of \( \delta \), the difference between these two situations is strictly a choice of parametrization.

Theorem 3.3.3: \( \Delta^n(\mu, I) \) is a strictly decreasing of \( \mu \).
Proof: Let $n=1$, then

$$\Delta^1(\mu, I) = -2\mu.$$ 

Suppose the theorem is true for $n$; we need to show it holds for $n+1$.

Let $\mu' < \mu''$, then

$$\Delta^n(\mu' + \frac{J_1}{I(I+J_1)}^{1/2} z, I+J_1) > \Delta^n(\mu'' + \frac{J_1}{I(I+J_1)}^{1/2} z, I+J_1)$$

and

$$\Delta^n(\mu' - \frac{J_2}{I(I+J_2)}^{1/2} z, I+J_2) > \Delta^n(\mu'' - \frac{J_2}{I(I+J_2)}^{1/2} z, I+J_2), \text{ for all } z.$$ 

which implies that

$$\Delta^n(\mu'+\frac{I}{I(I+J_1)}^{1/2} z, I+J_1)^+ > \Delta^n(\mu''+\frac{I}{I(I+J_1)}^{1/2} z, I+J_1)^+$$

and

$$\Delta^n(\mu'-z\frac{I}{I(I+J_2)}^{1/2} z, I+J_2)^- > \Delta^n(\mu''-(\frac{I}{I(I+J_2)}^{1/2} z, I+J_2)^-, \text{ for all } z.$$ 

A result similar to Theorem 2.6.5 shows that there exists a $z^*$ such that

$$\Delta^n(\mu'+\frac{I}{I(I+J_1)}^{1/2} z, I+J_1)^+ > \Delta^n(\mu''+\frac{I}{I(I+J_1)}^{1/2} z, I+J_1)^+$$

for all $z < z^*$.

Now,

$$\Delta^n(\mu', I) - \Delta^n(\mu'', I) = \left[ \Delta^n(\mu' + \frac{I}{I(I+J_1)}^{1/2} z, I+J_1)^+ ight.$$ 

$$\left. - \Delta^n(\mu'' + \frac{I}{I(I+J_1)}^{1/2} z, I+J_1)^+ d\phi(z) \right]$$

$$+ \left[ \Delta^n(\mu' - \frac{I}{I(I+J_2)}^{1/2} z, I+J_2)^- - \Delta^n(\mu'' - \frac{I}{I(I+J_2)}^{1/2} z, I+J_2)^- d\phi(z) \right] > 0.$$ 

Q.E.D.
The special case where $J_1 = J_2 = J$ is of special interest since the optimal strategy has a particularly simple structure. The optimal strategy is shown in this section to be completely determined by the sign of the posterior expected value of $\theta$. Therefore, for this special case, information has no role in the determination of the optimal strategy.

The following theorem is the singular prior analog of Lemma 2.4.2.

Theorem 3.4.1: $V_y^1(\mu, I) = V_x^1(-\mu, I)$, provided $J_1 = J_2 = J$.

Proof: Let $n=1$, then

\[ V_y^1(\mu, I) = -\mu \]

\[ = V_x^1(-\mu, I). \]

Suppose the theorem holds for $n$, we need to show that it holds for $n+1$.

First, notice that the singular prior analog of Lemma 2.3.1 implies that

\[ V_y^n(\mu, I) = V_x^n(-\mu, I), \]

since

\[ V_y^n(\mu, I) = \min(V_x^n(\mu, I), V_y^n(\mu, I)) \]

\[ = \min(V_y^n(-\mu, I), V_x^n(-\mu, I)) \]

\[ = V_x^n(-\mu, I). \quad (3.4.1) \]

Therefore, if we combine Equations 3.3.4, 3.3.5, and 3.4.1, we have that
\[ v_{y}^{n+1}(\mu, I) = -\mu + \int v^{n}(\mu - \left(\frac{I}{I(I+J)}\right)^{1/2}, I+J) d\phi(z) \]
\[ = -\mu + \int v^{n}(-\mu + \left(\frac{I}{I(I+J)}\right)^{1/2}, I+J) d\phi(z) \]
\[ = v_{x}^{n+1}(-\mu, I) . \]

Q.E.D.

The following corollary shows that \( \Delta_{y}(\mu, I) \) is an odd function in \( \mu \) for every fixed value of \( I \).

Corollary 3.4.2: \( \Delta_{y}(\mu, I) = -\Delta_{y}(-\mu, I) \), provided \( J_{1} = J_{2} = J \).

Proof:
\[ \Delta_{y}(\mu, I) = V_{y}^{n}(\mu, I) - V_{x}^{n}(\mu, I) \]
\[ = V_{x}^{n}(-\mu, I) - V_{y}^{n}(-\mu, I) \]
\[ = -\Delta_{y}(-\mu, I) . \]

Q.E.D.

The following corollary, when combined with Theorem 3.3.1, shows that the optimal strategy is completely determined by the sign of \( E(\theta|x, z) \).

Corollary 3.4.3: \( \Delta_{y}^{n}(\mu, I)^{\langle z \rangle} 0 \) if \( \mu^{\langle z \rangle} 0 \).

Proof:

Corollary 3.4.2 implies that
\[ \Delta_{y}^{n}(0, I) = 0 . \]

Now, Theorem 3.3.2 which shows that \( \Delta_{y}^{n}(\mu, I) \) is a strictly decreasing function of \( \mu \) implies that if \( \mu > 0 \) then
\[ \Delta_{y}^{n}(\mu, I) > \Delta_{y}^{n}(0, I) \]
\[ = 0 . \]
A similar inequality holds for the case where $\mu<0$.

Q.E.D.

Theorem 3.4.2 is significant since the optimal strategy can be characterized as follows if $J_1 = J_2 = J$:

i) sample source $X$ if $\mu^* < 0$

ii) sample source $Y$ if $\mu^* > 0$

iii) sample either source if $\mu^* = 0$

where

$$\mu^* = (I_1 + n_1n_2J)^{-1}(I_1\mu + J(n_1\overline{x} - n_2\overline{y}))$$

The singular prior case optimal strategy has a much simpler structure than that for the independent because information is gathered in the distribution of the entire mean vector $(\theta_1, \theta_2)$ regardless of the source sampled. For this reason, the singular prior case can be thought of as the normal analog of the Bernoulli Bandit in Feldman (1962). In both problems, there exists a dependence between the conditional distributions of sources.

The case where $J_1 = J_2 = J$ has a particularly simple optimal strategy since the same amount of information on the mean vector is obtained regardless of the source sampled. The situation of unequal variances has a more complicated optimal strategy because a slightly inferior source may be sampled if a large amount of information can be obtained by sampling that source.
4. BIBLIOGRAPHY


5. ACKNOWLEDGMENTS

First of all, I would like to thank Dr. H. T. David for his encouragement and unbounded optimism. Only his insightful suggestions and constant support enabled this work to be completed. Secondly, I would like to thank my wife, Marcey, for enduring the long hours devoted to completion of my degree and finally, I acknowledge my daughter, Jenny Kae, for providing a stabilizing influence during the hectic times of the past year.