1969

Minimax properties of likelihood ratio tests related to goodness of fit

Kenneth Stephan Mount
Iowa State University

Follow this and additional works at: https://lib.dr.iastate.edu/rtd
Part of the Statistics and Probability Commons

Recommended Citation
https://lib.dr.iastate.edu/rtd/3764

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.
MOUNT, Kenneth Stephan, 1942-
MINIMAX PROPERTIES OF LIKELIHOOD
RATIO TESTS RELATED TO GOODNESS
OF FIT.

Iowa State University, Ph.D., 1969
Statistics

University Microfilms, Inc., Ann Arbor, Michigan
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>I. INTRODUCTION AND LITERATURE REVIEW</td>
<td>1</td>
</tr>
<tr>
<td>II. THE SMALL SAMPLE PROBLEM</td>
<td>3</td>
</tr>
<tr>
<td>A. The Small Sample Problem with Fixed Probability of Type I Error</td>
<td>3</td>
</tr>
<tr>
<td>B. The Small Sample Problem with Arbitrary Probability of Type I Error</td>
<td>21</td>
</tr>
<tr>
<td>C. Stringency</td>
<td>25</td>
</tr>
<tr>
<td>III. MULTIVARIATE LARGE DEVIATIONS</td>
<td>29</td>
</tr>
<tr>
<td>A. Convexity of the ( \ln ) of a Moment Generating Function</td>
<td>29</td>
</tr>
<tr>
<td>for a Random Variable whose Distribution has Finite Support</td>
<td></td>
</tr>
<tr>
<td>B. Finite Minima of the Moment Generating Function</td>
<td>31</td>
</tr>
<tr>
<td>C. Bivariate Extension of Bahadur and Rao's Results</td>
<td>33</td>
</tr>
<tr>
<td>D. Moment Generating Functions and Multivariate Large Deviations</td>
<td>39</td>
</tr>
<tr>
<td>1. Multinomial large deviations</td>
<td>39</td>
</tr>
<tr>
<td>2. Multivariate large deviations for distributions with finite support</td>
<td>42</td>
</tr>
<tr>
<td>E. Multivariate Large Deviations for Exponential Families</td>
<td>45</td>
</tr>
<tr>
<td>IV. ASYMPTOTIC OPTIMALITY AND MINIMAXNESS</td>
<td>48</td>
</tr>
<tr>
<td>A. Asymptotic Optimality of a Certain Likelihood Ratio Procedure</td>
<td>48</td>
</tr>
<tr>
<td>B. Asymptotic Minimaxness in Exponential Families</td>
<td>70</td>
</tr>
<tr>
<td>V. BIBLIOGRAPHY</td>
<td>76</td>
</tr>
<tr>
<td>VI. ACKNOWLEDGEMENTS</td>
<td>77</td>
</tr>
</tbody>
</table>
I. INTRODUCTION AND LITERATURE REVIEW

In this thesis we consider optimality properties, both small sample and asymptotic, of likelihood ratio procedures for goodness of fit testing problems. In this connection we consider a testing problem to relate to goodness of fit when the distributions of null and alternative hypotheses differ in "shape".

In Chapter II, which contains small sample considerations, we introduce a certain power-slope criterion for goodness of fit testing problems. This in fact amounts to working with the familiar risk which is the sum of the probabilities of the errors of first and second kinds. Certain Neyman-Pearson procedures are shown to be minimax for this risk. Other considerations in Chapter II pertain to stringency and the use of a risk equal to the probability of the error of the second kind, the latter occurring when the error of the first kind is fixed.

In Chapter III we present a variety of multivariate large deviations results which depend in large part on the properties of cumulant generating functions (i.e., the natural logarithms of moment generating functions). These are used in Chapter IV for the asymptotic analysis.

Recent papers treating one-dimensional large deviations are those of Bahadur and Rao (1), Chernoff (3) and Sethuraman (12). Hoeffding (8) and Sanov (10) have treated the multinomial distribution. Sanov also considered the general, not necessarily multinomial, case which was recently quite thoroughly treated by Borovkov and Rogozin (2). Herr (7) extended Hoeffding's work to the case of the multivariate normal distribution. Efron and Truax (5) have treated exponential families.
Sections A and B of Chapter III discuss certain properties of bivariate cumulant generating functions. Based on A and B, Section C extends the first order portion of (1) to the bivariate case. In Section D we give an alternative derivation of the results of Borovkov and Rogozin (2) for distributions with finite support. Finally, Section E presents an aspect of large deviations theory for exponential families which is of special interest for the asymptotic minimax considerations of Section IV B.

In Section A of Chapter IV we consider arbitrary discrete null and alternative hypotheses. We derive a condition that insures asymptotic optimality and hence asymptotic minimaxness of a natural likelihood ratio procedure, \( \tilde{\delta} \), in a fairly strong sense. Here, as already motivated in Chapter II, the risk is the sum of the probabilities of Type I and Type II error. The approach is illustrated by two examples.

In Section B we consider exponential families and hypotheses that are disjoint subsets of the natural parameter space. We use the same risk as in Section A. It is shown that \( \tilde{\delta} \) is asymptotically minimax in a somewhat weaker sense than in Section A.

Prior work related to this is that of L. J. Savage (11) which indicates an asymptotic minimax property of maximum likelihood estimates. As to testing, Hoeffding, Herr and Efron and Truax have given related results concerning the asymptotic optimality of likelihood ratio tests in the multinomial, multivariate normal and exponential family cases respectively. All of these consider risk equal to the probability of the error of the second kind, with the behavior of the probability of the error of the first kind suitably restricted.
II. THE SMALL SAMPLE PROBLEM

A. The Small Sample Problem with Fixed Probability of Type I Error

A "good" goodness of fit test should be able to distinguish against alternatives whose shape is quite close to that of the null distributions. Stating this in a different manner, as the alternative becomes quite "distinct" from the null the power should increase rapidly. In fact, the greater the "slope" of the power "curve" as the alternative gets farther away from the null, the better. With this in mind we consider the hypothesis testing situation:

\[ H_0: f_0(x) \]
\[ H_A: (1-\theta_1-\theta_2)f_0(x) + \theta_1f_1(x) + \theta_2f_2(x) \quad \theta_1, \theta_2 \geq 0 \quad 0 < \theta_1 + \theta_2 \leq 1 \]

where \( f_0(x), f_1(x) \) and \( f_2(x) \) are three distinct densities. Here the power for a test \( \delta(x) \) is:

\[ \int [\delta((1-\theta_1-\theta_2)f_0 + \theta_1f_1 + \theta_2f_2)] \]

or, if we wish to consider a criterion of the probability of Type II error, we have:

\[ \beta_0(\theta_1, \theta_2) = 1 - \int [\delta((1-\theta_1-\theta_2)f_0 + \theta_1f_1 + \theta_2f_2)] \]

which is linear in \( \theta_1 \) and \( \theta_2 \) and has two slopes

\[ \beta_0^{(i)}(0,0) = \frac{\partial \beta_0}{\partial \theta_i} \bigg|_{\theta_1=\theta_2=0} = \int \delta f_0 - \int \delta f_i, \quad i = 1, 2 \]

the slope in any other direction from \( \theta_1 = \theta_2 = 0 \) being a linear combination of these two. If we take a conservative approach to this problem we
might seek to minimize the maximum of these two slopes. With the probability of Type I error fixed at \( \alpha \) (i.e., \( \int \delta f_0 = \alpha \)) this criterion becomes:

\[
\min \max \{ \beta_0'(0,0), \beta_0^2(0,0) \} = \min \max \{ \alpha - \int \delta f_1, \alpha - \int \delta f_2 \}. \tag{4}
\]

In this form the problem is somewhat intractable. However it may be rephrased in the terminology of game theory. Now we have two players, I and II, each player having a set of possible strategies. Player I's set is:

\[ A = \{0,1\} \]

with 0 indicating the choice of \( f_1 \) and 1 the choice of \( f_2 \). Player II's set is:

\[ B = \{ \delta(x): 0 \leq \delta(x) \leq 1 \ \forall x, \int \delta f_0 = \alpha \} \]

The payoff function of the game is:

\[ R'(\delta, \tau) = \alpha - \int \delta f_{\tau+1}, \quad \tau \in A \delta \in B \]

The sets \( A \) and \( B \) are the sets of pure strategies for the players. Mixed strategies are generated by defining probability measures on \( A \) and \( B \). Since \( A \) consists of two points the measure will be specified by the probability it gives to either of them. We denote a measure on the space \( A \) by \( \eta \).

Later (Theorem 2) we will show that Player II has a minimax strategy \( \delta_0 \), which is a member of \( B \) (i.e., a pure strategy). Thus it is not necessary to consider mixed strategies for Player II.

It is a property of games that their saddle points (i.e., the minimax strategies for Players I and II) are unchanged if a constant, say \( 1-\alpha \), is added to the payoff function. Thus the saddle point for the above game is
the same as that for the game with strategy spaces \( A \) and \( B \) and payoff function

\[
R(\delta, \xi) = 1 - \int_{A} \xi \, d\delta. \tag{5}
\]

Now we shall show that players I and II have minimax strategies.

From above, any measure \( \eta_i \) on \( A \) is uniquely determined by \( \eta_i(0) = 1-p_i \) say. Using this notation:

\[
R(\delta, \eta_i) = \int_{A} R(\delta, \xi) \, d\eta_i(\xi)
\]

\[
= R(\delta, 0)(1-p_i) + R(\delta, 1)p_i \tag{6}
\]

\[
= p_i [\int \delta f_1 - \int \delta f_2] + 1 - \int \delta f_1.
\]

Since \( 0 \leq p_i \leq 1 \forall i \) there exists a convergent subsequence \( \langle p_{ij} \rangle_{j=1}^{\infty} \) and a constant \( p_0 \) such that \( p_{ij} \to p_0 \) as \( j \to \infty \). Define \( \eta_0 \) by \( \eta_0(0) = 1 - p_0 \).

Then

\[
R(\delta, \eta_0) = p_0 [\int \delta f_1 - \int \delta f_2] + 1 + \int \delta f_1,
\]

\[
= \lim_{j} R(\delta, \eta_i) .
\]

Thus

\[
\lim \inf_{j} R(\delta, \eta_i) = R(\delta, \eta_0) \quad \delta \in B \tag{7}
\]

and \( A \) is weakly compact (13, p.53).

**Theorem 1** (13). If the space \( A \) is weakly compact, a minimax strategy for Player I exists.

Since knowing \( p_i \) is equivalent to knowing \( \eta_i \), we shall hereafter denote \( R(\delta, \eta_i) \) by \( R(\delta, p_i) \).
To show that Player II has a minimax strategy we apply the Weak Compactness Theorem (9, p. 354) with $\mu = \text{Lebesgue measure}$ $X = \mathbb{R}$ and $\mathcal{B} = \mathcal{B}$ the Borel sets. Since the class of sets $\{A_r\}_{r=1}^\infty$ where $A_r = (-\infty, r)$, $r$ rational, forms a countable number of generators of $\mathcal{B}$, the set of measurable functions $\delta(x)$ with $0 \leq \delta(x) \leq 1$ for all $x$ is compact with respect to weak convergence. i.e., for any function $g$ such that $\int g d\mu$ exists and for any sequence $<\delta_n>_{n=1}^\infty$ there exists a subsequence $<\delta_{n_i}>_{i=1}^\infty$ and a function $\delta$ such that:

$$\lim_{i \to \infty} \int \delta_n \cdot g d\mu = \int \delta g d\mu.$$ 

Thus

$$\lim_{i \to \infty} \int \delta_n \cdot f_j d\mu = \int \delta_f \cdot d\mu \quad j = 0, 1, 2$$

and

$$\lim_{i \to \infty} [1 - \int \delta_n \cdot f_j d\mu] = 1 - \int \delta_f \cdot d\mu \quad j = 0, 1, 2.$$  \hspace{1cm} (8)

For $j = 0$ $\int \delta_n \cdot f_0 d\mu = \alpha$, $i=1, 2, \ldots$, and

$$\lim_{i \to \infty} \int \delta_n \cdot f_0 d\mu = \int \delta_f \cdot d\mu.$$ 

Thus $\delta \in \mathcal{B}$.

**Theorem 2.** A minimax strategy for Player II exists.

**Pf.** Let $<\delta_n>_{n=1}^\infty$ be a sequence of members of $\mathcal{B}$ such that

$$\lim_{n \to \infty} \max_{\ell \in A} R(\delta_n, \ell) = \inf_{\delta \in \mathcal{B}} \max_{\ell \in A} R(\delta, \ell)$$  \hspace{1cm} (9)

where we use $\inf$ instead of $\min$ because $\mathcal{B}$ contains an infinite number of elements.
By the Weak Compactness Theorem there exists a subsequence $\delta_{i}^{n} \rightarrow \infty$ and a function $\delta_{0} \in B$ such that:

$$\lim_{i \rightarrow \infty} R(\delta_{i}^{n}, t) = R(\delta_{0}, t), \quad t \in A.$$ 

From above $\delta_{0} \in B$. Thus

$$\max R(\delta_{0}, t) \geq \inf \max R(\delta, t). \quad (10)$$

Also

$$\inf \max R(\delta, t) = \lim \max R(\delta_{i}^{n}, t) \quad (11)$$

since convergence of a sequence implies convergence of all its subsequences.

This quantity is $\geq \lim R(\delta_{i}^{n}, t)$ since $R(\delta_{i}^{n}, t) \leq \max R(\delta_{i}^{n}, t)$ for all $t$ in $B$ and since $\lim R(\delta_{i}^{n}, t)$ exists for all $t \in B$. This limit is equal to $R(\delta_{0}, t)$ i.e.,

$$\inf \max R(\delta, t) \geq R(\delta_{0}, t), \quad t \in B.$$ 

Since the left hand side of the inequality is independent of $t$

$$\inf \max R(\delta, t) \geq \max R(\delta_{0}, t). \quad (12)$$

From equations (10) and (11)

$$\inf \max R(\delta, t) = \max R(\delta_{0}, t), \quad t \in B.$$ 

i.e., $\delta_{0}$ is a minimax strategy for Player II. q.e.d.

We note that this proof essentially parallels the hint given for problem 1 (9, Chapt.8).
Since both players have optimal strategies the game has a saddle point
with coordinates $(\delta_0, p_0)$, say, and thus a value. Then $0 \leq p_0 \leq 1$ and $\delta_0 \in \mathcal{B}$.

We consider several cases:

1) $p_0 = 0$. Since $(\delta_0, p_0)$ is a saddle point $R(\delta_0, p_0)$ is minimum (inf.) in
its row and maximum in its column of the payoff matrix, i.e.

$$R(\delta_0, 0) = \inf_{\delta} R(\delta, 0)$$

$$= \inf_{\delta} [1 - \int \delta f_1]$$

$$= 1 - \sup_{\delta} \int \delta f_1$$

$$= 1 - \int_{\delta_{NP,1}} \int f_1$$

where $\delta_{NP,1}$ is the Neyman-Pearson test function for testing $H_0:f_0$ vs.
$H_A:f_1$ at the $\alpha$-level. In other words $(\delta_0, p_0) = (\delta_{NP,1}, 0)$ (under certain
conditions (9, p.65)).

2) $p_0 = 1$. Analogously to 1) we may show that (under the same conditions)
$\delta_0 = \delta_{NP,2}$, the Neyman-Pearson level $\alpha$ test function for testing $H_0:f_0$
vs. $H_A:f_2$.

3) $0 < p_0 < 1$. Now $R(\delta_0, p_0) = (1-p_0)R(\delta_0, 0) + p_0 R(\delta_0, 1)$ and using half
of the saddle point condition $R(\delta_0, p_0) = \max_{p_0} R(\delta_0, p) \geq R(\delta_0, t) \ t = 0,1$.
But $R(\delta_0, p_0)$ is a convex combination of $R(\delta_0, 0)$ and $R(\delta_0, 1)$. Thus

$$R(\delta_0, p_0) = R(\delta_0, 0) = R(\delta_0, 1).$$

In other words

$$\int \delta_0 f_1 = \int \delta_0 f_2$$

Also
\[ R(\delta_0, p_0) = \inf_{\delta} R(\delta, p_0) \]
\[ = \inf_{\delta} [(1-p_0)(1-\delta f_1) + p_0(1-\delta f_2)] \]
\[ = 1 - \sup_{\delta} \{ \int \delta [(1-p_0) f_1 + p_0 f_2] \} \quad (15) \]

Thus \( \delta_0 \) is the solution to the problem of maximizing \( \int [(1-p_0) f_1 + p_0 f_2] \)
subject to \( \int \delta f_1 = \int \delta f_2 \) and \( \int \delta f_0 = \alpha \). This problem has the same solution as
the problem of maximizing \( \int \delta (f_1 + f_2) \) subject to \( \int \delta f_1 = \int \delta f_2 \) and \( \int \delta f_0 = \alpha \).

\( \delta_0 \) can be found by using the Generalized Neyman-Pearson Lemma. If we
express the first constraint as \( \int \delta (f_1 - f_2) = 0 \) and apply the Lemma we obtain:

\[ \delta_0(x) = \begin{cases} 
1 & \text{when } f_1(x) + f_2(x) > K_1 f_0(x) + K_2[f_1(x) - f_2(x)] \\
0 & \text{when } f_1(x) + f_2(x) < K_1 f_0(x) + K_2[f_1(x) - f_2(x)] 
\end{cases} \]

\[ \delta_0(x) = \begin{cases} 
1 & \text{when } f_0(x) < f_1(x) \left( \frac{1-K_2}{K_1} \right) + f_2(x) \left( \frac{1+K_2}{K_1} \right) \\
0 & \text{when } f_0(x) > f_1(x) \left( \frac{1-K_2}{K_1} \right) + f_2(x) \left( \frac{1+K_2}{K_1} \right) 
\end{cases} \]

Since \( (\delta_0, p_0) \) is a saddle point we have from (15)

\[ R(\delta_0, p_0) = 1 - \sup_{\delta} \int \delta [(1-p_0) f_1 + p_0 f_2] \]
\[ = 1 - \int \delta_{NP, p_0} [(1-p_0) f_1 + p_0 f_2] \quad (16) \]

where \( \delta_{NP, p_0} \) is the most powerful level \( \alpha \) test function for testing \( H_0 : f_0 \)
vs. \( H_A : (1-p_0) f_1 + p_0 f_2 \) and is obtained by the Neyman-Pearson Lemma

\[ \delta_{NP, p_0}(x) = \begin{cases} 
1 & \text{when } f_0(x) < K(1-p_0) f_1(x) + K_0 f_2(x) \\
0 & \text{when } f_0(x) > K(1-p_0) f_1(x) + K_0 f_2(x) 
\end{cases} \]
Under certain conditions (9, p. 65) $\delta_{NP,P_0}$ is the only function which achieves this maximum. If the maximum is unique $\delta_0(x) = \delta_{NP,P_0}(x)$ a.e. and

$$\frac{1-K_2}{K_1} = K(1-p_0)$$

$$\frac{1+K_2}{K_1} = Kp_0$$

or

$$K = \frac{2}{K_1}$$

$$P_0 = \frac{1+K_2}{2}$$

Since $K > 0$ and $0 < p_0 < 1$, $K_1 > 0$ and $-1 < K_2 < 1$.

Having demonstrated that this game is completely determined we may give an algorithm for its solution:

1) Find $\delta_{NP,1}$. If $\int \delta_{NP,1} f_1 \leq \int \delta_{NP,1} f_2$ then $(\delta_0, p_0) = (\delta_{NP,1}, 0)$.

2) If $\int \delta_{NP,1} f_1 > \int \delta_{NP,1} f_2$ find $\delta_{NP,2}$. If $\int \delta_{NP,2} f_2 \leq \int \delta_{NP,2} f_1$, then $(\delta_0, p_0) = (\delta_{NP,2}, 1)$.

3) If $\int \delta_{NP,1} f_1 > \int \delta_{NP,1} f_2$ and $\int \delta_{NP,2} f_2 > \int \delta_{NP,2} f_1$ find $\delta_3$, the Generalized Neyman-Pearson solution to the problem of maximizing $\int \delta(f_1+f_2)$ subject to $\int \delta f_0 = \alpha$ and $\int \delta f_1 = \int \delta f_2$. Then $(\delta_0, p_0) = (\delta_3, \frac{1+K_2}{2})$.

Thus the solution is obtained by applying either the "simple" or the generalized form of the Neyman-Pearson Lemma and the minimax test function is a "simple" Neyman-Pearson test function, i.e., it can be obtained by applying the "simple" Neyman-Pearson Lemma to get a test of $H_0 : f_0$ vs. the appropriate (simple) alternative hypothesis.
Having enumerated the several possible saddle points we can give necessary and sufficient conditions for each of them to be the saddle point:

1) \((\delta_o, P_o) = (\delta_{NP,1}, 0) \iff R(\delta_{NP,1}, 0) \geq R(\delta_{NP,1}, 1) \text{ and } R(\delta_{NP,2}, 1) < R(\delta_{NP,2}, 0)\)

2) \((\delta_o, P_o) = (\delta_{NP,2}, 1) \iff R(\delta_{NP,1}, 0) < R(\delta_{NP,1}, 1) \text{ and } R(\delta_{NP,2}, 1) \geq R(\delta_{NP,2}, 0)\)

3) \((\delta_o, P_o) = (\delta_{NP,2}, 0) \iff R(\delta_{NP,1}, 0) < R(\delta_{NP,1}, 1) \text{ and } R(\delta_{NP,2}, 1) < R(\delta_{NP,2}, 0)\)

4) \((\delta_o, P_o) = (\delta_{NP,1}, 0) \text{ or } (\delta_{NP,2}, 1) \iff R(\delta_{NP,1}, 0) \geq R(\delta_{NP,1}, 1) \text{ and } R(\delta_{NP,2}, 1) \geq R(\delta_{NP,2}, 0)\)

in this case \(R(\delta_{NP,1}, 0) > R(\delta_{NP,1}, 1) \geq R(\delta_{NP,2}, 1) \geq R(\delta_{NP,2}, 0)\)

\(R(\delta_{NP,1}, 0) = R(\delta_{NP,2}, 1)\)

i.e., both saddle points have the same payoff; we know this must be true if a game has more than one saddle point.

Now let us generalize the problem and test

\[H_0: X \sim f_0(x)\]

vs.

\[H_A: X \sim (1 - \sum_{i=1}^K \theta_i f_0(x) + \sum_{i=1}^K \theta_i f_i(x) \quad \theta_i \geq 0 \quad 0 < \sum \theta_i \leq 1\] (17)

where for any test function \(\delta, \int \delta f_0 = \alpha\) and \(f_0, f_1, \ldots, f_K\) are \(K + 1\) distinct densities. The probability of Type II error is:

\[\beta_\delta(\theta_1, \ldots, \theta_K) = 1 - \int \delta \left[ (1 - \sum \theta_i) f_0 + \sum \theta_i f_i \right] \]
We want to find the test that maximizes \(\min[\beta_0^1(0,\ldots,0), \beta_0^2(0,\ldots,0), \ldots, \beta_0^K(0,\ldots,0)]\) where

\[
\beta_0^i(0,\ldots,0) = \frac{\mathcal{A}_{\theta_0}(\theta_1,\ldots,\theta_K)}{\mathcal{A}_{\theta_i}} |_{\theta_1=\ldots=\theta_K=0} \\
= \alpha - \int \delta f_i.
\]

Thus

\[
\max \min[\beta_0^1(0,\ldots,0), \ldots, \beta_0^K(0,\ldots,0)] = \max \min[\alpha - \int \delta f_1, \ldots, \alpha - \int \delta f_K].
\]

Again we rephrase the problem in game theory terms. The strategy space for Player II is unchanged while the space for Player I becomes:

\[A^I = \{0,1,\ldots,K-1\} \]

The payoff function is

\[R'(\delta,t) = \alpha - \int \delta f_{t+1}. \quad t \in A^I, \delta \in B.\]

The above comments about mixed strategies for Player II are again valid. Mixed strategies for Player I are measures defined over \(K\) points (i.e., \(\eta(0),\eta(1),\ldots,\eta(K-1)\)). The proof given above (Theorem 2) applies here and shows that Player II has a (pure) minimax strategy. We will not change the minimax strategies for either player if we add a constant \((1-\alpha)\) to the payoff function. The new payoff function is:

\[R(\delta,t) = 1 - \int \delta f_{t+1}. \quad t \in A^I, \delta \in B.\]

The set of Player I's mixed strategies is

\[\{\eta: \eta(i) \geq 0 \quad i = 0,1,\ldots,K-1; \sum_{i=0}^{K-1} \eta(i) = 1\}\]

and
Consider a sequence \( \eta_j^{\infty} \) of measures on \( A' \) or equivalently a sequence of points \( p_{(j)}^{(1)}, p_{(j)}^{(2)}, \ldots, p_{(j)}^{(j)} \) in \( E_K \) where \( p_{(j)}^{(i)} = \eta_j(i) \). As above we shall denote \( R(\delta, \eta) \) by \( R(\delta, p) \) where \( p = (p_0, \ldots, p_{K-1}) \) and \( p_i = \eta(i) \). Each of these points are in the set

\[
C = \{ p: p = (p_0, \ldots, p_{K-1}); p_i \geq 0 \text{ for } i = 0, 1, \ldots, K-1; \sum p_i = 1 \}
\]

which is compact (if we use the usual Euclidean metric as the distance function). Thus there exists a subsequence \( p_{(j)}^{(m)} \) and a point \( p^{(o)} \in C \) such that \( p_{(j)}^{(m)} \rightarrow p^{(o)} \) as \( m \rightarrow \infty \). Thus \( p_{(j)}^{(m)} \rightarrow p^{(o)} \) and

\[
\lim_{m \rightarrow \infty} R(\delta, p_{(j)}^{(m)}) = \lim_{m \rightarrow \infty} \sum_{\ell=0}^{K-1} R(\delta, t)p_{\ell}^{(m)} = \sum_{\ell=0}^{K-1} R(\delta, t)p_{\ell}^{(o)} = R(\delta, p)^{(o)}
\]

i.e., \( C \) is weakly compact and by Theorem 1 a minimax strategy for Player I exists.

Since both players have minimax strategies the game has a saddle point \( (\delta_0, \eta_0) \) and thus a value. There are several possibilities for the saddle point:

1) \( \eta_0(t) = 1 \quad 0 \leq t \leq K-1 \)
\[ \eta_O(j) = 0 \quad j \neq \ell \]

Now \[ R(\delta_O, \eta_O) = R(\delta_O, \ell) \]
\[ = \inf_{\delta} R(\delta, \ell) \]
\[ = \inf_{\delta} \left[ 1 - \int_{\ell+1}^\infty \delta f_{\ell+1} \right] \]
\[ = 1 - \int_{\infty}^{\infty} \delta_{\eta_NP_{\ell+1}} f_{\ell+1} \quad (20) \]

where \( \delta_{\eta_NP_{\ell+1}} \) is the Neyman-Pearson level \( \alpha \) test function for testing \( f_0 \) vs. \( f_{\ell+1} \).

2) \( \eta_O \) has positive values at exactly two integers say \( r \) and \( s \). Let \( \eta_O(r) = p \) then \( \eta_O(s) = 1 - p \), and \( \eta_O(j) = 0 \quad j \neq r, s \).

\[ R(\delta_O, \eta_O) = \inf_{\delta} R(\delta, \eta_O) \]
\[ = \inf_{\delta} \left[ p R(\delta, r) + (1-p) R(\delta, s) \right] \]
\[ = \inf_{\delta} \left[ 1 - p \int_{r+1}^{\infty} \delta f_{r+1} - (1-p) \int_{s+1}^{\infty} \delta f_{s+1} \right] \]
\[ = 1 - \sup_{\delta} \int \delta \left[ p f_{r+1} + (1-p) f_{s+1} \right] \]

Thus \( \delta_O \) is the Neyman Pearson level \( \alpha \) test function for testing \( H_0 : f_0 \) vs. \( H_A : p f_{r+1} + (1-p) f_{s+1} \). But since \( p \) is unknown \( H_A \) is unknown and we can't find \( \delta_O \). However since \( R(\delta_O, \eta_O) \) is a saddle point

\[ R(\delta_O, \eta_O) = \sup_{\eta} R(\delta_O, \eta) \quad (21) \]

Let \( \eta_L \) be defined by \( \eta_L(j) = \delta_{j \ell} \quad \ell = 0, 1, \ldots, K \). By (21)

\[ R(\delta_O, \eta_L) \geq R(\delta_O, \eta^{(r)}) \]
But \( R(\delta_0, \eta_0) \) is a convex combination of \( R(\delta_0, \eta^r) \) and \( R(\delta_0, \eta^s) \).

Thus:
\[
R(\delta_0, \eta_0) = R(\delta_0, \eta^r) = R(\delta_0, \eta^s) \quad (22)
\]

From (22)
\[
R(\delta_0, \eta^r) = R(\delta_0, \eta^s)
\]
or
\[
\int \delta_0 f_{r+1} = \int \delta_0 f_{s+1}
\]

We conclude that \( \delta_0 \) is a solution to the problem of maximizing
\[
\int \delta_0 (pf_{r+1} + (1-p)f_{s+1}) \text{ subject to the constraints } \int \delta f_0 = \alpha \text{ and } \int \delta f_{r+1} = \int \delta f_{s+1}.
\]
This is equivalent to the problem of maximizing \( \int \delta (f_{r+1} + f_{s+1}) \) subject to \( \int \delta f_0 = \alpha \) and \( \int \delta f_{r+1} = \int \delta f_{s+1} \) which may be solved by applying the Generalized Neyman-Pearson Lemma; the value of \( p \) can be found as above.

3) If \( \eta_0 \) has positive values at exactly \( t \) integers \( (3 \leq t \leq K) \) the analysis is analogous to that above. We have more constraints, \( t \), when applying the Generalized Neyman-Pearson Lemma but we also have more unknown values of \( \eta_0(i), t-1 \).

An algorithm for the solution of this game is:

1) Find \( \delta_{NP,2} \). If \( R(\delta_{NP,2}, \eta) = \max R(\delta_{NP,2}, \eta) \) then the saddle point is \( (\delta_{NP,2}, \eta(0)) \). This maximum is rather easy to find since the payoff \( R(\delta, \eta) \) is, for any \( \delta \) and \( \eta \), just a convex linear combination of \( R(\delta, \eta^t) \) for \( t = 0, \ldots, K-1 \). Thus \( \max R(\delta, \eta) = \max R(\delta, \eta^t) \) for any \( \delta \) and in particular for \( \delta_{NP,2} \).
2) If \( \max R(\delta_{NP}, 1, \eta) > R(\delta_{NP}, 1, 0) \) find \( \delta_{NP, 2} \). If \( R(\delta_{NP}, 2, 1) = \max \eta R(\delta_{NP}, 2, \eta) \) then \( (\delta_0, \eta_0) = (\delta_{NP, 2}, \eta^{(1)}) \). If \( R(\delta_{NP}, 2, 1) < \max \eta R(\delta_{NP}, 2, \eta) \) find \( \delta_{NP, 3} \). Continue in this manner until the saddle point is found or all \( \delta_{NP, i} \), \( i=1, 2, \ldots, K \) are exhausted. (Of course as soon as we find a saddle point we stop.)

3) If none of the \( \delta_{NP, i} \) produce a saddle point take all pairs of integers \( i \) and \( j \) (\( i < j \)) between 1 and \( K \), find the Generalized Neyman-Pearson solution for each pair, (i.e., \( \max\delta(f_i + f_j) \) subject to \( \int \delta f_i = \int \delta f_j \) and \( \int \delta f_0 = \alpha \)) and check if this solution, \( \delta_{ij} \) say, is one coordinate of a saddle point. (From above all we need do is show \( R(\delta_{ij}^*, i) = R(\delta_{ij}^*, j) \geq R(\delta_{ij}^*, t) \) \( t \neq i, j \)). If it is, we can find \( \eta_0 \) by finding \( \eta_0(i) = p \). As above \( p = \frac{1+K}{2} \). Then \( \eta_0(j) = 1-p, \eta_0(t) = 0 \) \( t \neq i, j \).

4) If we still have not found a saddle point we next consider all sets of three (distinct) integers in the same way, then all sets of four, etc., up to the set of all \( K \) integers at the same time.

There may be more than one saddle point but the payoff at each will be the same. Thus the order in which the above steps are performed is actually irrelevant. They are listed in order of increasing complexity.

We might hope that in the solution of the above game we do not have to resort to the Generalized Neyman-Pearson Lemma; in other words we hope that:

\[
\frac{\max_i R(\delta_{NP, i+1}, i)}{\inf_{\delta} \max_i R(\delta, i)} = 1
\]

(23)

i.e., the saddle point occurs at a value of \( \eta \) which is not a mixture of the two pure strategies. However the following example shows that this is generally not the case.
Suppose that the densities in the above game are:

\[
f_0(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}
\]

\[
f_1(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(x-1)^2}
\]

\[
f_2(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(x+1)^2}
\]

and that \( \alpha = .05 \). Then

\[
\delta_{NP,1}(x) = \begin{cases} 
  1 & x > 1.64 \\
  0 & x \leq 1.64 
\end{cases}
\]

\[
\delta_{NP,2}(x) = \begin{cases} 
  1 & x <-1.64 \\
  0 & x \geq -1.64 
\end{cases}
\]

The only possibilities for pure saddle points are \((\delta_{NP,1}, 0)\) and \((\delta_{NP,2}, 1)\). If \((\delta_{NP,1}, 0)\) is a saddle point then

\[
R(\delta_{NP,1}, 0) = \max_{\eta} R(\delta_{NP,1}, \eta) \\
\geq R(\delta_{NP,1}, 1) \quad . \quad (24)
\]

But

\[
R(\delta_{NP,1}, 0) = 1 - \int_{\delta_{NP,1}} f_1 \\
= 1 - P_{f_1} \{X > 1.64\} \\
= 1 - P_{f_0} \{X > .64\} \\
= .739
\]
and
\[
R(\delta_{NP,1}, 1) = \int_0^\infty \delta_{NP,1}^f \cdot 2
\]
\[
= 1 - P_{f_2} \{X > 1.64\}
\]
\[
= 1 - P_{f_0} \{X > 2.64\}
\]
\[
= .99 .
\]

Thus \((\delta_{NP,1}, 0)\) is not a saddle point. If \((\delta_{NP,2}, 1)\) is a saddle point then
\[
R(\delta_{NP,2}, 1) = \max_\eta R(\delta_{NP,2}, \eta)
\]
\[
\geq R(\delta_{NP,2}, 1)
\]

But
\[
R(\delta_{NP,2}, 1) = 1 - \int_0^\infty \delta_{NP,2}^f \cdot 2
\]
\[
= 1 - P_{f_2} \{X < -1.64\}
\]
\[
= 1 - P_{f_0} \{X < -0.64\}
\]
\[
= .739
\]

and
\[
R(\delta_{NP,2}, 1) = 1 - \int_0^\infty \delta_{NP,2}^f \cdot 1
\]
\[
= 1 - P_{f_1} \{X < -1.64\}
\]
\[
= .99 .
\]

Therefore this game does not have a pure saddle point. To find the saddle point we apply the Generalized Neyman-Pearson Lemma with the constraints
\[
\int_0 \delta f_1 = \int_0 \delta f_2 \text{ and } \int_0 \delta f_0 = \alpha
\]
Each expression below is equivalent to the one following it:

\[ f_0(x) < f_1(x) \left( \frac{1-K_2}{K_1} \right) + f_2(x) \left( \frac{1+K_2}{K_1} \right) \]

\[ \frac{x^2}{2} < e^{-\frac{(x-1)^2}{2} \left( \frac{1-K_2}{K_1} \right)} + e^{-\frac{(x+1)^2}{2} \left( \frac{1+K_2}{K_1} \right)} \]

\[ 1 < e^{\frac{2x-1}{2} \left( \frac{1-K_2}{K_1} \right)} + e^{\frac{-2x-1}{2} \left( \frac{1+K_2}{K_1} \right)} \]

\[ 1 < e^x \left[ e^{-\frac{1}{2} \left( \frac{1-K_2}{K_1} \right)} \right] + e^{-x} \left[ e^{-\frac{1}{2} \left( \frac{1+K_2}{K_1} \right)} \right] \quad (26) \]

If \( e^{\frac{2x-1}{2} \left( \frac{1-K_2}{K_1} \right)} > 0 \) and \( e^{\frac{-2x-1}{2} \left( \frac{1+K_2}{K_1} \right)} > 0 \) equation (26) defines a region of the form

\[ \{ x : x < C_1 < 0 \text{ or } x > C_2 > 0 \} \]

By trying various values for \( C_1 \) and \( C_2 \) we can find such a region which satisfies the above constraints:

\[ \{ x : x < -1.96 \text{ or } x > 1.96 \} \]

To find \( p \), we note that at \( x = -1.96 \) and 1.96 we have

\[ 1 = e^x \left[ e^{-\frac{1}{2} \left( \frac{1-K_2}{K_1} \right)} \right] + e^{-x} \left[ e^{-\frac{1}{2} \left( \frac{1+K_2}{K_1} \right)} \right] \]

i.e.

\[ 1 = 7.1 \left[ e^{-\frac{1}{2} \left( \frac{1-K_2}{K_1} \right)} \right] + 0.14 \left[ e^{-\frac{1}{2} \left( \frac{1+K_2}{K_1} \right)} \right] \]
Thus
\[ K = 1.4 \left[ e^{-\frac{1}{2} \left( \frac{1-K_2}{K_1} \right)} \right] + 7.1 \left[ e^{-\frac{1}{2} \left( \frac{1+K_2}{K_1} \right)} \right]. \]

Thus
\[ e^{-\frac{1}{2} \left( \frac{1-K_2}{K_1} \right)} = \begin{pmatrix} 1 & 1.4 \\ 7.1 & 1.4 \\ 1.4 & 7.1 \end{pmatrix} = 0.138 \]

and
\[ e^{-\frac{1}{2} \left( \frac{1+K_2}{K_1} \right)} = e^{-\frac{1}{2} \left( \frac{1-K_2}{K_1} \right)} \]

or
\[ \frac{1-K_2}{K_1} = \frac{1+K_2}{K_1} \]

from which
\[ K_2 = 0. \]

Thus
\[ P_0 = \frac{1+K_2}{2} = 0.5 \]

and the optimal strategy \( \eta_0 \) for Player 1 is \( \eta_0(0) = \eta_0(1) = 0.5 \). The value of this game is
\[ R(\delta_0, \eta_0) = \frac{1}{2} R(\delta_0, 0) + \frac{1}{2} R(\delta_0, 1) \]
\[ = 0.83 \]
\[ > \max R(\delta_{NP, i+1, i}) \]
B. The Small Sample Problem with Arbitrary Probability of Type I Error

If instead of just \( X \) we had a sample of size \( n \) the tests would be the analogous ones, e.g., \( \delta_{\text{NP},1}(X_1, \ldots, X_n) = 1 \iff \prod_{i=1}^{n} \frac{f_1(X_i)}{f_0(X_i)} > C_{n, \alpha} \). Having found a minimax test (of a certain type) for samples of size \( n \) it is natural to ask about its asymptotic properties (e.g., consistency). There are two difficulties we encounter immediately in considering asymptotic properties:

1) For a fixed \( n \) the minimax test is one of three possible tests. As \( n \) changes the minimax test may change from say \( \delta_{\text{NP},1} \) to \( \delta_{\text{NP},2} \) or to the Generalized Neyman-Pearson test. Thus we may be considering the asymptotic behavior of a test that oscillates among three tests as \( n \) changes.

2) Even if we could show that the minimax test did not oscillate as \( n \) changed, \( \delta_o = \delta_{\text{NP},1} \) for all \( n \) say, the minimax test would still depend on a constant (e.g., \( \delta_{\text{NP},1}(X_1, \ldots, X_n) = 1 \iff \prod_{i=1}^{n} \frac{f_1(X_i)}{f_0(X_i)} > C_{n, \alpha} \)) which depended on \( n \) and \( \alpha \). The asymptotic behavior of \( \delta_o \) would depend upon the asymptotic behavior of this constant which might be difficult to ascertain.

To avoid at least part of this difficulty we drop the restriction that \( \int \delta f_0 = \alpha \) and consider the above game with variable probability of Type I error. Since both Type I and Type II errors are now varying we will need a new risk function. We have noted that adding a constant to the risk function does not change the location (or existence) of minimax strategies for either player. Thus in the previous game we would have obtained the same \( \delta_o \) and \( \eta_o \) had we used

\[
R(\delta, \eta) = P[\text{Type II error}]
\]
or

\[ R(\delta, \eta) = P[\text{Type II error}] + \alpha \]

\[ = P[\text{Type II error}] + P[\text{Type I error}] \]

It is with this in mind that we define the risk for the variable \( \alpha \) game to be:

\[ R_1(\delta, \ell) = P[\text{Type II error}] + P[\text{Type I error}] \]

\[ = 1 - \int \delta f_{\ell+1} + \int \delta f_0 \quad \ell = 0, 1 \]  \hfill (27)

The strategy spaces for the variable \( \alpha \) game are \( A \) and \( B'' = \{ \delta: 0 \leq \delta \leq 1 \} \).

Mixed strategies are as defined above. Similarly to the above no mixed strategies on the space \( B'' \) are needed. Also

\[ R_1(\delta, \eta_i) = \int_A R_1(\delta, \ell) \, d\eta_i(\ell) \]

\[ = R_1(\delta, 0)(1-p_i) + R_1(\delta, 1)p_i \]

\[ = (1-p_i)[1 - \int \delta f_1 + \int \delta f_0] + p_i [1 - \int \delta f_2 + \int \delta f_0] \]

\[ = p_i[\int \delta f_1 - \int \delta f_2] + 1 + \int \delta f_1 + \int \delta f_0 \]

Proceeding as in the fixed \( \alpha \) game we can show that the space \( A \) is weakly compact and Player I has a minimax strategy.

To show that a minimax strategy for Player II exists we can apply the proof we used in the fixed-\( \alpha \) game deleting equation (8) and in place of it noting that \( \lim \int \delta \eta_i f_j = \int \delta \omega f_j \) \( j = 0, 1, 2 \) implies that
\[ \lim_{i} \left[ 1 - \sum_{i} \delta_{i} f_{i} + \sum_{i} \delta_{i} f_{0} \right] = 1 - \sum_{i} \delta_{0} f_{i} + \sum_{i} \delta_{0} f_{0} \quad \ell = 1, 2 \]

\[ \text{i.e.,} \quad \lim_{i} R(\delta_{i}, t) = R(\delta_{0}, t) \quad \ell = 0, 1. \]

Since both players have optimal strategies the game has a saddle point, say \((p_{o}, \delta_{o})\), and thus a value. We consider several cases:

1) \(p_{o} = 0\). Now

\[ R_{1}(\delta_{o}, p_{o}) = \inf_{\delta} R_{1}(\delta, p_{o}) \]

\[ = \inf_{\delta} R_{1}(\delta, 0) \]

\[ = \inf_{\delta} \left[ 1 - \sum \delta f_{1} + \sum \delta f_{0} \right] \]

\[ = 1 - \sup_{\delta} \sum \delta (f_{1} - f_{0}) \]

from which we conclude that

\[ \delta_{o}(x) = \begin{cases} 
1 & \text{when } f_{1}(x) - f_{0}(x) > 0 \\
0 & \text{otherwise}
\end{cases} \]

or

\[ \delta_{o}(x) = \begin{cases} 
1 & \text{when } \frac{f_{1}(x)}{f_{0}(x)} > 1 \\
\frac{f_{1}(x)}{f_{0}(x)} & \text{when } \frac{f_{1}(x)}{f_{0}(x)} \leq 1
\end{cases} \]

(28)

i.e., the test function is a likelihood ratio function.

2) \(p_{o} = 1\). Analogously to 1) we may show that

\[ \delta_{o}(x) = \begin{cases} 
1 & \text{when } \frac{f_{2}(x)}{f_{0}(x)} > 1 \\
\frac{f_{2}(x)}{f_{0}(x)} & \text{when } \frac{f_{2}(x)}{f_{0}(x)} \leq 1
\end{cases} \]

(29)
3) $0 < p_o < 1$. Now $R_1(\delta_o, p_o) = p_o R_1(\delta_o, 0) + (1-p_o) R_1(\delta_o, 1)$ and as before (fixed-$\alpha$ case)

$$R_1(\delta_o, p_o) = R_1(\delta_o, 0) = R_1(\delta_o, 1).$$
(30)

From equation (30)

$$1 - \int \delta_o f_1 + \int \delta_o f_0 = 1 - \int \delta_o f_2 + \int \delta_o f_0$$

or

$$\int \delta_o f_1 = \int \delta_o f_2.$$

Since $(\delta_o, p_o)$ is a saddle point $R_1(\delta_o, p_o) = \inf \limits_\delta R(\delta, p_o)$ subject to

$$\int \delta f_1 = \int \delta f_2 \text{ or}$$

$$R_1(\delta_o, p_o) = \inf \limits_\delta R_1(\delta, p_o)$$

$$= \inf \limits_\delta \{p_o[1 - \int \delta f_1 + \int \delta f_0] + (1-p_o)[1 - \int \delta f_2 + \int \delta f_0]\}$$

$$= \inf \limits_\delta \{1 + \int \delta f_0 - p_o \int \delta f_1 - (1-p_o) \int \delta f_2\}$$

$$= \inf \limits_\delta \{1 + \int \delta f_0 - p_o \int \delta f_1\}$$

$$= \inf \limits_\delta \{1 + \int \delta f_0 - \int \delta f_1\}$$

$$= 1 - \sup \limits_\delta \int \delta(f_1 - f_0)$$.

Thus by reasoning as in the fixed-$\alpha$ case $\delta_o$ is the solution to the problem of maximizing $\int \delta(f_1 - f_0)$ subject to $\int \delta f_1 = \int \delta f_2$ or equivalently the problem of maximizing $\int \delta(f_1 + f_2 - 2f_0)$ subject to $\int \delta f_1 = \int \delta f_2$. By applying the Generalized Neyman-Pearson Lemma
\[
\delta_o(x) = \begin{cases} 
1 & \text{when } f_1(x) + f_2(x) > 2f_0(x) > K(f_1(x) - f_2(x)) \\
0 & \text{when } f_1(x) + f_2(x) - 2f_0(x) \leq K(f_1(x) - f_2(x))
\end{cases}
\]

or

\[
\delta_o(x) = \begin{cases} 
1 & \text{when } f_1(x) \left(\frac{1-K}{2}\right) + f_2(x) \left(\frac{1+K}{2}\right) > f_0(x) \\
0 & \text{when } f_1(x) \left(\frac{1-K}{2}\right) + f_2(x) \left(\frac{1+K}{2}\right) \leq f_0(x)
\end{cases}
\]

C. Stringency

Having found a minimax test for \(H_0\) vs. \(H_A\) we are now interested in finding a most stringent test. Specifically, we wish to find a level \(\alpha\) test function \(\delta\) which minimizes

\[
\max \left\{ \left[ \frac{\partial \beta_1}{\partial \alpha} - \frac{\partial \beta_{\delta N P 1}}{\partial \alpha} \right]_{\theta_1 = \theta_2 = 0}, \left[ \frac{\partial \beta_2}{\partial \alpha} - \frac{\partial \beta_{\delta N P 2}}{\partial \alpha} \right]_{\theta_1 = \theta_2 = 0} \right\}
\]

where \(\delta_{NP1}\) and \(\delta_{NP2}\) have been defined previously. Evaluating derivatives we may write this as

\[
\min \max \{\int \delta_{NP1} f_1 - \int \delta f_1, \int \delta_{NP2} f_2 - \int \delta f_2\}
\]

If we put this in game theory terms A and B will be as above and the pay-off (or risk) function will be:

\[
R_2(\delta, i) = \int \delta_{NP3, i+1} f_{i+1} - \int \delta f_{i+1}, \quad i = 0, 1
\]

Mixed strategies are as above; none is needed for Player II and they are defined as above for Player I. The proofs that Players I and II have minimax strategies proceed similarly to those in Section A of Chapter II. Thus our new game has a saddle point (and a value). Denoting the saddle
point by \((\delta_o, p_o)\) we have several possibilities:

1) \(p_o = 0\). Now \(R_2(\delta_o, P_o) = \inf_\delta R_2(\delta_o, 0)\)

\[
= \inf_\delta [\int_{\delta_{NP_1}} f_1 - \int_{\delta f_1}]
\]

\(= 0\) when \(\delta_o = \delta_{NP_1}\)

2) \(p_o = 1\). Proceeding as in 1) \(\delta_o = \delta_{NP_2}\).

3) \(0 < p_o < 1\). By reasoning as in Section A of Chapter II we may show that

\[
R_2(\delta_o, P_o) = R_2(\delta_o, 0) = R_2(\delta_o, 1)
\]

i.e.,

\[
\int_{\delta_{NP_1}} f_1 - \int_{\delta f_1} = \int_{\delta_{NP_2}} f_2 - \int_{\delta f_2}
\]

(34)

Also

\[
R_2(\delta_o, P_o) = \inf_\delta R_2(\delta, P_o)
\]

\[
= \inf_\delta [(1-p_o)[\int_{\delta_{NP_1}} f_1 - \int_{\delta f_1} + p_o[\int_{\delta_{NP_2}} f_2 - \int_{\delta f_2}]]
\]

\[
= (1-p_o)\int_{\delta_{NP_1}} f_1 + p_o\int_{\delta_{NP_2}} f_2 - \sup_\delta [\int_{\delta} [(1-p_o)f_1 + p_o f_2]]
\]

Lemma 1: The problem of maximizing \(\int_{\delta} [(1-p_o)f_1 + p_o f_2]\) subject to (34) and \(\int_{\delta} f_0 = \alpha\) is equivalent to the problem of maximizing \(\int_{\delta} (f_1 + f_2)\)

subject to (34), and \(\int_{\delta} f_0 = \alpha\).

Pf.: Let us rewrite (34) as

\[
\int_{\delta} f_1 = \int_{\delta} f_2 + \int_{\delta_{NP_1}} f_1 - \int_{\delta_{NP_2}} f_2
\]

Given that (34) holds for some \(\delta\)
\[ J_6(f_1 + f_2) = 2J_6f_2 + J_6N_{B1}f_1 - J_6N_{B2}f_2 \]

Then, since \( J_6N_{B1}f_1 - J_6N_{B2}f_2 \) is constant, maximizing \( J_6(f_1 + f_2) \) subject to (34) and \( J_6f_0 = \alpha \) is equivalent to maximizing \( J_6f_2 \) subject to (34) and \( J_6f_0 = \alpha \). Again given (34) and \( J_6f_0 = \alpha \)

\[ J_6[(1-p_o)f_1 + p_0f_2] = (1-p_o)J_6f_2 + (1-p_o)[J_6N_{B1}f_1 - J_6N_{B2}f_2] \]

+ \( p_0J_6f_2 \)

\[ = J_6f_2 + (1-p_o)[J_6N_{B1}f_1 - J_6N_{B2}f_2] \]

Since \( p_0 \) is fixed, maximizing \( J_6[(1-p_o)f_1 + p_0f_2] \) subject to (34) and \( J_6f_0 = \alpha \) is equivalent to maximizing \( J_6f_2 \) subject to the same two constraints. Both problems are equivalent to the same problem and hence to each other. q.e.d.

We can find \( \max J_6(f_1 + f_2) \) such that (34) and \( J_6f_0 = \alpha \) hold by using the Generalized Neyman-Pearson Lemma. Call the resulting test function \( \delta_3 \).

Then

\[
\delta_3(x) = \begin{cases} 
1 & \text{when } f_1(x) + f_2(x) > K_1(f_2(x) - f_1(x)) + K_2 f_0(x) \\
0 & \text{when } f_1(x) + f_2(x) < K_1(f_2(x) - f_1(x)) + K_2 f_0(x) 
\end{cases}
\]

i.e.,

\[
\delta_3(x) = \begin{cases} 
1 & \text{when } f_0(x) < f_1(x) \left(\frac{1+K_1}{K_2}\right) + f_2(x) \left(\frac{1-K_1}{K_2}\right) \\
0 & \text{when } f_0(x) > f_1(x) \left(\frac{1+K_1}{K_2}\right) + f_2(x) \left(\frac{1-K_1}{K_2}\right) 
\end{cases}
\]

Similarly to Section A of Chapter II we may, under certain conditions (9, p.65), equate a.e. \( \delta_3 \) to \( \delta_3' \), the Neyman-Pearson test function for testing
If we do this and

$$\delta_3'(x) = \begin{cases} 
1 & \text{when } f_0(x) < K(1-p_o)f_1(x) + Kp_0f_2(x) \\
0 & \text{when } f_0(x) > K(1-p_o)f_1(x) + Kp_0f_2(x) 
\end{cases}$$

we obtain

$$K = \frac{2}{K_2}$$

$$p_o = \frac{1-K_1}{2}$$

Since $K > 0$ and $0 < p_o < 1$, $K_2 > 0$ and $-1 < K_1 < 1$.

An algorithm for the solution of this game is:

1) Find $\delta_{NB1}$. If $R_2(\delta_{NB1}, 0) = 0 = \sup_p R_2(\delta_{NB1}, p)$ then $(\delta_o, p_o) = (\delta_{NB1}, 0)$.

But

$$R_2(\delta_{NP1}, p) = (1-p) R_2(\delta_{NP1}, 0) + p R_2(\delta_{NP1}, 1)$$

$$= p [\int \delta_{NP2} f_2 - \int \delta_{NB1} f_2]$$

which will usually (9, p.65) be positive for all $p \neq 0$. Thus $(\delta_o, p_o) = (\delta_{NB1}, 0)$ is unlikely.

2) If $R_2(\delta_{NP1}, 0) < \max_p R_2(\delta_{NP1}, p)$ find $\delta_{NB2}$. Again $(\delta_o, p_o) = (\delta_{NB2}, 1)$ is unlikely.

3) If $R_2(\delta_{NB1}, 0) < \max_p R_2(\delta_{NP1}, p)$ and $R_2(\delta_{NB1}, 1) < \max_p R_2(\delta_{NP2}, p)$ find $\delta_3$. Then

$$(\delta_o, p_o) = (\delta_3, \frac{1-K_1}{2}).$$
III. MULTIVARIATE LARGE DEVIATIONS

A. Convexity of the $\ln$ of a Moment Generating Function for a Random Variable whose Distribution has Finite Support

It will be useful for some of the asymptotic considerations in Chapter IV, to state and derive certain facts concerning bivariate large deviations. All of these can be extended to $K$ dimensions in a straightforward manner. The first fact concerns the strict convexity of the log of a bivariate moment generating function $\phi$, a fact to be used below which is demonstrated under certain regularity conditions. These essentially concern the legitimacy of differentiations under the integral sign.

Denote by $S$ the region of $E_2$ on which $\phi$ is finite. Let

$$\psi(t_1, t_2) = \ln \phi(t_1, t_2)$$

have second order derivatives $\psi^{20}, \psi^{02}$ and $\psi^{11}$ obtainable by differentiation under the integral sign. Then the strict convexity of $\psi$ is equivalent to the positive definiteness of the matrix

$$\begin{pmatrix}
\psi^{20}(t_1, t_2) & \psi^{11}(t_1, t_2) \\
\psi^{11}(t_1, t_2) & \psi^{02}(t_1, t_2)
\end{pmatrix}$$

for all $(t_1, t_2)$.

The $\psi^{ij}$ are computed as follows:

$$\psi^{20} = \frac{(20)}{(00)} - \left[ \frac{(10)}{(00)} \right]^2$$

$$\psi^{02} = \frac{(02)}{(00)} - \left[ \frac{(01)}{(00)} \right]^2$$
\[ \psi_{11} = \begin{pmatrix} 11 \\ 00 \end{pmatrix} - \begin{pmatrix} 10 \\ 00 \end{pmatrix} \begin{pmatrix} 01 \\ 00 \end{pmatrix} \]

where
\[ (ij) = E[x^i y^j e^{t_1 x + t_2 y}] \]

Suppose \((X, Y)\) has density \(p(x, y)\). Then
\[ \psi_{20} = \left( \frac{\iint x^2 e^{t_1 x + t_2 y} p(x, y) \, dx \, dy}{\iint e^{t_1 x + t_2 y} p(x, y) \, dx \, dy} \right)^2 - \left( \frac{\iint e^{t_1 x + t_2 y} p(x, y) \, dx \, dy}{\iint e^{t_1 x + t_2 y} p(x, y) \, dx \, dy} \right)^2 \]
\[ = \iint x^2 \left( \frac{e^{t_1 x + t_2 y} p(x, y)}{\iint e^{t_1 u + t_2 v} p(u, v) \, du \, dv} \right) \, dx \, dy - \left[ \iint x \left( \frac{e^{t_1 x + t_2 y} p(x, y)}{\iint e^{t_1 u + t_2 v} p(u, v) \, du \, dv} \right) \, dx \, dy \right]^2 \]
\[ = \iint x^2 p^* (x, y) \, dx \, dy - \left[ \iint x p^* (x, y) \, dx \, dy \right]^2 \]
\[ = \text{var}(x) \]

where the density \(p^*(x, y)\) is defined by:
\[ p^*(x, y) = \frac{e^{t_1 x + t_2 y} p(x, y)}{\iint e^{t_1 u + t_2 v} p(u, v) \, du \, dv} . \]

Similarly we can show that with respect to this density:
\[ \psi_{02} = \text{var}(y) \]
\[ \psi_{11} = \text{cov}(x, y) . \]

Thus equation (1) represents a covariance matrix. Finally, since \(p(x, y)\) and \(p^*(x, y)\) are singular together, \((X, Y)\) having a non-singular distribution
implies that the matrix in equation (1) is non-singular (and thus positive
definite) for every \((t_1, t_2)\). (We shall say that \((U, V)\) with distribution \(G\)
has a singular distribution if no point is in the interior of the convex
hull of the support of \(G\). This means that \(U\) and \(V\) are perfectly corre-
lated.)

B. Finite Minima of the Moment Generating Function

In order to derive our large deviations results we need some more
properties of the moment generating function, namely we need to know when
it has a finite minimum. We consider a random variable \((X, Y)\) whose distri-
bution has support \(\Omega_1\). Let \(C(\Omega_1)\) be the convex hull of \(\Omega_1\) and let \(\Omega_i\) be
the support of \((\bar{X}_i, \bar{Y}_i)\). Let \(F\) be the distribution function of \((X, Y)\).

**Lemma 1:** If \(C(\Omega_1)\) contains a neighborhood in each quadrant of \(\mathbb{R}^2\) then there
exists an integer \(m\) such that \(\Omega_m\) contains a neighborhood in each quadrant.

**Pf.:** \(C(\Omega_1)\) contains a neighborhood in each quadrant \(\Rightarrow C(\Omega_1)\) contains a
sphere, say \(S\), in quadrant \(Q\) centered at \(x = \exists x_1, x_2, x_3 \in \Omega_1\) and \(\theta_i \geq 0\)
\(i=1, 2, 3\) \(\Rightarrow \sum_i \theta_i = 1\) and \(\sum_i \theta_i x_i = x = \exists \delta > 0\), rational numbers \(\alpha_i\)
\(i=1, 2, 3\) and \(N_i\) a neighborhood of \(x_i\). \(i=1, 2, 3\) \(\Rightarrow \max_i |\theta_i - \alpha_i| < \delta\) and \(y_i \in N_i\)
\(i=1, 2, 3\) \(\Rightarrow \sum_i \alpha_i y_i \in S\). Since the \(\alpha_i\) are rational \(\exists m \Rightarrow \alpha_i = \frac{K_i}{m}\) \(i=1, 2, 3\).

Let \(m = K_1 + K_2 + K_3\) and suppose \(x_1, \ldots, x_{K_1} \in N_1; x_{K_1+1}, \ldots, x_{K_1+K_2} \in N_2\) and
\(x_{K_1+K_2+1}, \ldots, x_m \in N_3\). Then, denoting the mean of the \(i^{th}\) group \(x_{K_1}^i\),
\(\bar{x}_{K_1}^i \in N_1\) \(i=1, 2, 3\) which means that \(\bar{x}_m = \frac{1}{m} \sum_i K_i \bar{x}_{K_1}^i \in S\). Since the probability
of the \(x_i\)'s being in the indicated neighborhoods is positive, the proba-
bility that \(\bar{x}_m \in S\) is also positive. q.e.d.

We can now state the conditions for the moment generating function to have
a finite minimum.
Lemma 2: If \( \Omega_1 \) contains a neighborhood in each quadrant then the moment generating function of \( (X, Y) (\phi(t_1, t_2)) \) is minimized at finite \( (t_1, t_2) \).

**Pf.:** \( \forall \varepsilon > 0 \) define \( Q_{1, \varepsilon} = \{ (x, y): x \geq \varepsilon, y \geq \varepsilon \} \)

\[ Q_{2, \varepsilon} = \{ (x, y): x \leq -\varepsilon, y \geq \varepsilon \} \]

\[ Q_{3, \varepsilon} = \{ (x, y): x \leq -\varepsilon, y \leq -\varepsilon \} \]

\[ Q_{4, \varepsilon} = \{ (x, y): x \geq \varepsilon, y \leq -\varepsilon \} \]

Then since \( \Omega_1 \) contains a neighborhood in each quadrant \( \exists \varepsilon > 0 \quad \Omega_1 \cap Q_{i, \varepsilon} \neq \emptyset \)

\( i = 1, 2, 3, 4 \), i.e., \( P\{Q_{1, \varepsilon}\} > 0 \quad i = 1, 2, 3, 4 \). Then if \( t_1 > 0, t_2 > 0 \)

\[ \phi(t_1, t_2) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} t_1 x + t_2 y dF(x, y) \]

\[ = \int_{Q_{1, \varepsilon}} \int_{-Q_{1, \varepsilon}} t_1 x + t_2 y dF(x, y) \]

\[ \geq \int_{Q_{1, \varepsilon}} \int_{Q_{1, \varepsilon}} (t_1 + t_2) \min(x, y) dF(x, y) \]

\[ \geq e^{(t_1 + t_2) \varepsilon} \int_{Q_{1, \varepsilon}} \int_{Q_{1, \varepsilon}} dF(x, y) \]

\[ = e^{(t_1 + t_2) \varepsilon} P\{Q_{1, \varepsilon}\} \]

and \( e^{(t_1 + t_2) \varepsilon} P\{Q_{1, \varepsilon}\} \to \infty \) as \( t_1 \to \infty \) or \( t_2 \to \infty \).

Similarly using \( Q_{2, \varepsilon}, Q_{3, \varepsilon}, \) and \( Q_{4, \varepsilon} \) we can show that \( \phi(t_1, t_2) \to \infty \) as \( t_1 \to \pm \infty \) or \( t_2 \to \pm \infty \). Also by using the above result and the analogous ones for \( Q_{2, \varepsilon}, Q_{3, \varepsilon}, \) and \( Q_{4, \varepsilon} \) it is clear that \( \forall \delta > 0 \exists r \sqrt{t_1^2 + t_2^2} > r \Rightarrow \)
\( \varphi(t_1, t_2) > 1 + \delta \). Since we know \( \varphi(0,0) = 1 \), inf \( \varphi(t_1, t_2) \) must occur inside this circle. Furthermore since \( \varphi(t_1, t_2) \) is a continuous function and since \[ \{(t_1, t_2) : \sqrt{t_1^2 + t_2^2} \leq r \} \] is compact \( \varphi(t_1, t_2) \) not only has an inf it has a minimum in this circle. q.e.d.

**Lemma 3:** If \( C(\Omega_i) \) contains a neighborhood in each quadrant then \( \varphi(t_1, t_2) \) is minimized at a finite \( (T_1, T_2) \).

**Pf.:** By Lemmas 1 and 23 the moment generating function of \((X_m, Y_m)\), say \( \psi(t_1, t_2) \), is minimized at finite \( (\omega_1, \omega_2) \). But

\[
\psi(t_1, t_2) = \left[ \varphi\left( \frac{t_1}{m}, \frac{t_2}{m} \right) \right]^m.
\]

Thus if \( \psi \) is minimized at \( (\omega_1, \omega_2) \) \( \varphi \) is minimized at \( T_1 = \frac{\omega_1}{m}, T_2 = \frac{\omega_2}{m} \). q.e.d.

**C. Bivariate Extension of Bahadur and Rao's Results**

One method for evaluating the bivariate first quadrant probabilities that we shall encounter in Chapter IV is to generalize the results of Bahadur and Rao (1). We consider a random variable \((X, Y)\) and use the notation of Section B.

**Theorem 1** If \( C(\Omega_i) \) contains a neighborhood in each quadrant then \( \varphi(t_1, t_2) \) is minimized at a finite \( (T_1, T_2) \). If \( T_1 > 0, T_2 > 0 \) (a necessary condition for this is that \( (E(X), E(Y)) \) not be in the first quadrant) then:

\[
P[\tilde{X}_n \geq 0, \tilde{Y}_n \geq 0] = \rho n^{\alpha_1 \alpha_2} \int_0^\infty \int_0^\infty \exp\left[-n^{1/2}[\alpha_1 v_1 + \alpha_2 v_2]\right] \left[ H_n(v_1, v_2) - H_n(v_1, 0) - H_n(0, v_2) - H_n(0, 0) \right] dv_1 dv_2 \quad (2)
\]

where

\[
\rho = \varphi(T_1, T_2)
\]
\[ \alpha_i = \sigma_i T_i \quad i = 1, 2 \]

\[ \sigma_i = \text{var}(Z_i) \quad i = 1, 2 \]

\[ G(z_1, z_2) = \frac{1}{\rho} \int_{-\infty}^{z_2} \int_{-\infty}^{z_1} e^{i x y} \, dF(x, y) \]

is the distribution function of \((Z_1, Z_2)\)

\[ H_n(v_1, v_2) = \text{distribution function of} \quad \left( \frac{k^{1/2}}{n^{1/2} \sigma_1}, \frac{k^{1/2}}{n^{1/2} \sigma_2} \right). \]

We also assume that \(G(z_1, z_2)\) has a density \(g(z_1, z_2)\).

Pf.: Proceeding analogously to (1) we introduce the variables \(Z_1\) and \(Z_2\). By mimicking the proof of Lemma 1 of (1) we can show that

\[ E(Z_i) = 0 \]

\[ 0 < \text{var}(Z_i) < \infty \quad i = 1, 2 \]

and

\[ \text{var}(Z_i) = \frac{1}{\phi(t_1, t_2)} \left. \frac{\partial^2 \varphi(t_1, t_2)}{\partial t_1^2} \right|_{t_1 = T_1, t_2 = T_2} \]

\[ \text{cov}(Z_1, Z_2) = \frac{1}{\phi(t_1, t_2)} \left. \frac{\partial^2 \varphi(t_1, t_2)}{\partial t_1 \partial t_2} \right|_{t_1 = T_1, t_2 = T_2}. \]

Define \(\alpha_i = \sigma_i T_i \quad i = i, 2\).

Let \((Z_{11}, Z_{12}, \ldots, Z_{1n})\), \((Z_{21}, Z_{22}, \ldots, Z_{2n})\) be a sequence of i.i.d.r.v. and let

\[ U_{1n} = \frac{Z_{11} + \ldots + Z_{1n}}{n^{1/2} \sigma_1} \]

\[ U_{2n} = \frac{Z_{21} + \ldots + Z_{2n}}{n^{1/2} \sigma_2} \]
with \( H_n(v_1, v_2) \) being the distribution function of \( \frac{U_{1n}}{U_{2n}} \), i.e.,

\[
H_n(v_1, v_2) = P(U_{1n} < v_1, U_{2n} < v_2)
\]

Then

\[
P(\bar{y}_n \geq 0, \bar{v}_n \geq 0) = P(\frac{1}{n} \gamma_1^1 + \ldots + \frac{1}{n} \gamma_n^n \geq \gamma_0^n) = P(\frac{X_1^1}{\gamma_1^1} + \ldots + \frac{X_n^n}{\gamma_n^n} \geq \gamma_0^n) = P(\gamma_1 + \ldots + \gamma_n \in Q_1) \quad \text{where} \quad Q_1 = \{ (\gamma_1^1, \ldots, \gamma_n^n) \}
\]

is the first quadrant

\[
= \int \ldots \int \int \ldots \int \int \int e^{\sum_{i=1}^{n} \sum_{j=1}^{n} z_{ij} g_{ij}} dG(z_{11}, \ldots, z_{nn})
\]

From the definition of \( G(z_1, z_2) \):

\[
dG(z_1, z_2) = \frac{1}{\rho} e^{T_1 z_1 + T_2 z_2} dF(z_1, z_2)
\]

Thus changing coordinates from \( y_i = (\gamma_i^1) \) to \( g_i = (z_{2i}) \) we have

\[
P(\bar{y}_n \geq 0, \bar{v}_n \geq 0) = P(\sum_{i=1}^{n} \sum_{j=1}^{n} z_{1i} g_{ij} + \sum_{i=1}^{n} \sum_{j=1}^{n} z_{2i} g_{ij} \geq \sum_{i=1}^{n} \sum_{j=1}^{n} z_{2i} g_{ij})
\]

Now let us transform to \( \bar{y}_1 = \sum_{i=1}^{n} g_i, \bar{y}_2 = \sum_{i=1}^{n} g_i, \ldots, \bar{y}_n = g_n \). Here \(|J|=1\) and
Since the distribution function $G$ has density $g$

\[ P\{\bar{X}_n \geq 0, \bar{Y}_n \geq 0\} = \rho^n \int_{\mathbb{R}^n} \cdots \int_{\mathbb{R}^n} e^{-\frac{1}{2} \sum \bar{Y}_i} g(\bar{Y}_1 - \frac{n}{2} \bar{Y}_1) g(\bar{Y}_2) \cdots g(\bar{Y}_n) d\bar{Y}_1 \cdots d\bar{Y}_n. \]

Consider the part of the integral which depends on $u_2$

\[ \int_{\mathbb{R}^2} g(\bar{Y}_1 - \frac{n}{3} \bar{Y}_1 - u_2) g(u_2) du_2. \]

This is just the convolution of $\bar{Y}_1 - \frac{n}{3} \bar{Y}_1$ with $u_2$ which yields the density of $U_1 - \frac{n}{3} U_1$. Next we consider the part of the integral which depends on $u_3$

\[ \int_{\mathbb{R}^2} g(\bar{Y}_1 - \frac{n}{4} \bar{Y}_1 - u_3) g(u_3) du_3. \]

Again this is the convolution of $U_1 - \frac{n}{4} U_1$ with $u_3$ and yields the density of $U_1 - \frac{n}{4} U_1$. Continuing in this manner and denoting the density of $U_1$ by $h_0(U_1)$ we have:

\[ P\{\bar{X}_n \geq 0, \bar{Y}_n \geq 0\} = \rho^n \int_{\mathbb{R}^n} e^{-\frac{1}{2} \sum \bar{Y}_i} h_0(U_1) dU_1. \]

Now let us perform the transformation

\[ v_1 = \frac{U_{11}}{\sqrt{n} \sigma_1}, \quad v_2 = \frac{U_{12}}{\sqrt{n} \sigma_2}. \]

Then
\[ P\{X_n \geq 0, Y_n \geq 0\} = \rho \iint_{\mathbb{R}^2} e^{-T^I Y_1(\psi)} dH_O(\psi) \quad \psi \in \mathbb{R}^2 \]

But
\[ dH_O(\psi) = P\{n^{1/2} \sigma_1 Y_1 < U_{11} < \frac{1}{n^{1/2}} \sigma_1 (Y_1 + \Delta Y_1) ; \]
\[ \frac{1}{n^{1/2}} \sigma_2 Y_2 < U_{12} < \frac{1}{n^{1/2}} \sigma_2 (Y_2 + \Delta Y_2) \}

since \[ U_{ii} = n^{1/2} \sigma_i Y_i \quad i = 1, 2. \] This can also be written
\[ = P\{v_1 < \frac{U_{11}}{n^{1/2} \sigma_1} < v_1 + \Delta Y_1 ; v_2 < \frac{U_{12}}{n^{1/2} \sigma_2} < v_2 + \Delta Y_2 \}
\]
\[ = P\{v_1 < v_1 + \Delta Y_1 ; v_2 < v_2 + \Delta Y_2 \}
\]
\[ = dH_n(\psi) \]

and
\[ T^I Y_1(\psi) = \frac{T_1}{n^{1/2} \sigma_1} Y_1 + \frac{T_2}{n^{1/2} \sigma_2} Y_2
\]
\[ = n^{1/2} [\alpha_1 Y_1 + \alpha_2 Y_2]
\]
\[ = n^{1/2} \alpha^I \psi . \]

Thus
\[ P\{X_n \geq 0, Y_n \geq 0\} = \rho \iint_{\mathbb{R}^2} e^{-n^{1/2} \alpha^I \psi} dH_n(\psi) \quad \psi \in \mathbb{R}^2
\]
\[ = \rho \int_0^\infty \int_0^\infty e^{-n^{1/2} (\alpha_1 Y_1 + \alpha_2 Y_2)} dH_n(Y_1, Y_2) \quad \psi \in \mathbb{R}^2
\]

Since G has a density g, H_n has a density h_n and
\[ P\{X_n \geq 0, Y_n \geq 0\} = \rho \iint_{\mathbb{R}^2} e^{-n^{1/2} (\alpha_1 Y_1 + \alpha_2 Y_2)} h_n(Y_1, Y_2) dY_1 dY_2 \quad \psi \in \mathbb{R}^2 \]
Now by integrating by parts twice, once with respect to each variable, we obtain equation (2). q.e.d.

Having obtained this somewhat cumbersome formula we can show that the dominant part of it is $p^n$.

**Lemma 4** \( P\{ \bar{X}_n \geq 0, \bar{Y}_n \geq 0 \} \leq p^n \forall n. \)

**Pf.** Similar to the proof of the analogous fact in one dimension (1). q.e.d.

**Lemma 5** \( \forall p_0 < p, P\{ \bar{X}_n \geq 0, \bar{Y}_n \geq 0 \} \geq p^n_0 \) for all sufficiently large \( n \).

**Pf.** We first note that \( \lim_{n \to \infty} H_n(v_1, v_2) = \bar{\Phi}_2(v_1, v_2) \) where \( \bar{\Phi}_2(v_1, v_2) \) is the distribution function of a bivariate normal distribution with mean \( (0, 0) \) and covariance matrix the identity matrix evaluated at \( (v_1, v_2) \). Let \( \varepsilon_1 > 0 \) and \( \varepsilon_2 > 0 \). Then if we let \( l_n \) denote everything on the right hand side of (2) except \( p^n \) we have:

\[
\begin{align*}
  l_n &\geq n^{1/2} \alpha_1 \alpha_2 \int_{\varepsilon_1}^\infty \int_{\varepsilon_2}^\infty \exp\left[-n^{1/2}[\alpha_1 v_1 + \alpha_2 v_2]\right] \left[ H_n(v_1, v_2) - H_n(0, v_2) - H_n(0, 0) + H_n(0, 0) \right] \, dv_1 \, dv_2 \\
  &\geq n^{1/2} \alpha_1 \alpha_2 \int_{\varepsilon_1}^\infty \int_{\varepsilon_2}^\infty \exp\left[-n^{1/2}[\alpha_1 v_1 + \alpha_2 v_2]\right] \left[ H_n(\varepsilon_1, \varepsilon_2) - H_n(0, \varepsilon_2) - H_n(0, 0) + H_n(0, 0) \right] \, dv_1 \, dv_2 \\
  &= n^{1/2}[H_n(\varepsilon_1, \varepsilon_2) - H_n(\varepsilon_1, 0) - H_n(0, \varepsilon_2) + H_n(0, 0)] \exp\left[-n^{1/2}[\alpha_1 \varepsilon_1 + \alpha_2 \varepsilon_2]\right].
\end{align*}
\]

Thus

\[
\log l_n \geq \log n^{1/2} + \log[H_n(\varepsilon_1, \varepsilon_2) - H_n(\varepsilon_1, 0) - H_n(0, \varepsilon_2) + H_n(0, 0)] - n^{1/2}[\alpha_1 \varepsilon_1 + \alpha_2 \varepsilon_2]
\]

and, dividing by \( n^{1/2} \).
\[ n^{-1/2} \log l_n \geq n^{-1/2} \log n^{1/2} + \log \left( \frac{H_n(\xi_1, \xi_2) - H_n(0, 0)}{H_n(\xi_1, 0) - H_n(0, \xi_2) + H_n(0, 0)} \right) \]

\[ - (\alpha_1 \xi_1 + \alpha_2 \xi_2) \quad (3) \]

As \( n \to \infty \) the first and second terms on the right hand side \( \to 0 \) and are negligible with respect to the third term on the right hand side. Therefore we can approximate (3) by

\[ n^{-1/2} \log l_n \geq - (\alpha_1 \xi_1 + \alpha_2 \xi_2) \]

Thus

\[ \lim \inf \{ n^{-1/2} \log l_n \} \geq - (\alpha_1 \xi_1 + \alpha_2 \xi_2) \]

But \( l_n \) can be written

\[ l_n = \int_e^{\xi_1, y_1} h_o(\xi_1) \, d\xi_1 \]

\[ \leq \int_e^{\xi_1, y_1} h_o(\xi_1) \, d\xi_1 \]

\[ \leq 1 \]

and since \( \xi_1 \) and \( \xi_2 \) are arbitrary it follows that \( n^{-1/2} \log l_n = o(1) \).

Hence by Theorem 1

\[ n^{-1} \log P[\bar{X}_n \geq 0, \bar{Y}_n \geq 0] = \log \rho + o(1) \]

and this is equivalent to the desired result. q.e.d.

D. Moment Generating Functions and Multivariate Large Deviations

1. Multinomial large deviations

Consider a \( k \)-nomial random variable \( \mathfrak{p} = (n_1, \ldots, n_k) \) with parameters \( N \) and \( \mathfrak{p} = (p_1, \ldots, p_k), \ p > 0 \):
\[ P[n_1 = v_1, \ldots, n_k = v_k] = \frac{N!}{\prod \nu_i!} \prod \nu_i \]

where \( v_i, i = 1, \ldots, k \) are any non-negative integers whose sum is \( N \). Consider the region

\[ S_N = \Sigma \cap R \cap G_N \]

where \( \Sigma \) is the open \( k \) simplex \( \{x: \sum x_i = 1, x_i > 0, i = 1, \ldots, k\} \) with \( x = (x_1, \ldots, x_k) \), \( G_N \) is the set of \( k \)-vectors whose coordinates are multiples of \( N^{-1} \) and \( R \) is the region \( \{x: \sum a_i x_i \geq 0; \sum b_i x_i = 0\} \).

Let \( \tilde{R} \) be the boundary of \( R \) in \( E_k \) and define

\[ \rho = \inf_{\xi \in \Sigma \cap R} \sum x_i \ln \frac{x_i}{\nu_i} \]

Then according to Sanov (10), if \( p \in \mathbb{R} \),

\[ P[N, S_N, p] \equiv P[N^p] \mathbb{P} \sim \rho \]  \hspace{1cm} (4)

where \( P(c) \sim \rho \) is taken to mean that \( P(c) \) can be approximated by a quantity whose dominant term is \( e^{-N\rho} \). When this occurs we shall say that \( P(c) \) is of magnitude \( \rho \). In other words, for any \( \epsilon > 0 \)

\[ \frac{P[N, S_N, p]}{e^{-(-\rho + \epsilon)N}} \to 0 \]

and

\[ \frac{e^{-(-\rho - \epsilon)N}}{P[N, S_N, p]} \to 0 \]

as \( n \) approaches \( \infty \).

A more explicit computation of \( \rho \) proceeds as follows: Define

\[ R_{1, \lambda} = \{x: \sum a_i x_i = 0; \sum b_i x_i = \lambda\}, \quad R_{2, \lambda} = \{x: \sum a_i x_i = \lambda; \sum b_i x_i = 0\} \]
B = \sup_{x \in \Sigma} \sum_{i=1}^{k} b_i x_i \text{ subject to } \sum_{i=1}^{k} a_i x_i = 0 \text{ and } A = \sup_{x \in \Sigma} \sum_{i=1}^{k} a_i x_i \text{ subject to } \sum_{i=1}^{k} b_i x_i = 0. \text{ Then}

\Sigma \cap R = \left[ \bigcup_{0 \leq \lambda < B} R_{1,\lambda} \bigcup_{0 \leq \lambda < A} R_{2,\lambda} \right]

and

\rho = \min_{0 \leq \lambda < B} \inf_{c_{1,\lambda}} \inf_{c_{2,\lambda}} \{ c_{1,\lambda}, c_{2,\lambda} \} \quad (5)

where

\begin{align*}
  c_{j,\lambda} &= \inf_{x \in R_{j,\lambda}} \sum_{i=1}^{k} x_i \ln\left( \frac{x_i}{p_i} \right) \quad j = 1, 2.
\end{align*}

Concerning the computation of the \( c_{j,\lambda} \), \( j = 1, 2 \) a straightforward extension of the argument on page 218 of (10) shows that

\begin{equation}
  c_{1,\lambda} = \lambda \sigma_2(\lambda) - \psi(\sigma_1(\lambda), \sigma_2(\lambda)) \quad (6)
\end{equation}

where \( \psi(t_1, t_2) \) is the natural logarithm of the moment generating function of the bivariate distribution \( F \) which assigns probabilities \( p_i \) to the points \( (a_i, b_i) \). Also \( (\sigma_1(\lambda), \sigma_2(\lambda)) \) is the solution of

\begin{equation}
  0 = \frac{\partial}{\partial t_1} \psi(t_1, t_2) \bigg|_{(\sigma_1(\lambda), \sigma_2(\lambda))} \quad (7)
\end{equation}

\begin{equation}
  \lambda = \frac{\partial}{\partial t_2} \psi(t_1, t_2) \bigg|_{(\sigma_1(\lambda), \sigma_2(\lambda))} \quad .
\end{equation}

Similarly

\begin{equation}
  c_{2,\lambda} = \lambda T_1(\lambda) - \psi(T_1(\lambda), T_2(\lambda)) \quad (8)
\end{equation}

where \( (T_1(\lambda), T_2(\lambda)) \) is the solution of
Next we verify the existence and uniqueness of the solutions of (7) and (9); specifically, say, of (9).

Concerning existence, \((T_1(\lambda), T_2(\lambda))\) is in fact the minimizing point of the moment generating function of the bivariate distribution which assigns probability \(p_i\) to the point \((a_i - \lambda, b_i)\). But according to Section B of Chapter III, such a finite minimizing point exists if the origin is in the interior of the convex hull of the support of that distribution; i.e., if \((\lambda, 0)\) is in the interior, \(C(\bar{\Omega})\), of the convex hull, \(C(\Omega)\), of the support, \(\Omega\), of \(F\). Uniqueness is guaranteed by the convexity argument of Section A of Chapter III if the origin is in \(C(\bar{\Omega})\). This is true because the Jacobian of the map in (9) from \((T_1, T_2)\) to \(\frac{\partial \psi}{\partial t_1}(T_1, T_2), \frac{\partial \psi}{\partial t_2}(T_1, T_2)\) is the inverse of matrix \(\psi^{ij}\) \(i, j = 1, 2\) of Section A of Chapter III (which we have shown to be positive definite for all \((T_1, T_2)\)) and all the assumptions of that section are satisfied in this case.

Finally, if the origin is in \(C(\bar{\Omega})\) then so are all points \((\lambda, 0)\) with \(\lambda \in [0, A)\) and all points \((0, \lambda)\) with \(\lambda \in [0, B)\). Hence, in summary, if the origin is in \(C(\bar{\Omega})\) then (4) holds with \(p\) given by (5), (6) and (8).

2. Multivariate large deviations for distributions with finite support

Borovkov and Rogozin (2) have given a very general theory of multivariate large deviations featuring the moment generating function. In this section we show how a specialization of their results to the case of finite
support follows by suitable approximation from the multinomial case.

Let \((X,Y)\) have a bivariate distribution with finite support \(\Omega\). Let 
\(\psi(t_1, t_2)\) be the natural logarithm of the moment generating function of this 
distribution. Let 
\(A = \sup \{y : (0, y) \in C(\bar{\Omega}) \} \) and let 
\(B = \sup \{x : (x, 0) \in C(\bar{\Omega}) \}. \)
(Note that by an application of a continuous version of the Duality Theorem 
of linear programming \(A\) is also \(\inf_s g_1(s)\) where \(g_1(s)\) is the upper envelope 
of the set of lines \(b - ax\) with \((a,b)\in \Omega\). Similarly \(B\) is also \(\inf_s g_2(s)\) 
where \(g_2(s)\) is the upper envelope of the set of lines \(a - bx\) with \((a,b)\in \Omega\).)

Also define 
\[
\rho = \min \left\{ \inf_{0 \leq \lambda < A} \left[ \lambda T_2(\lambda) - \psi(T_1(\lambda), T_2(\lambda)) \right], \inf_{0 \leq \lambda < B} \left[ \lambda \sigma_2(\lambda) - \psi(\sigma_1(\lambda), \sigma_2(\lambda)) \right] \right\}
\]

where \(T_i(\lambda)\) and \(\sigma_i(\lambda)\) \(i = 1,2\) are defined as in Section D.1 of Chapter III.

Then if the origin is in \(C(\bar{\Omega})\) we now show by an approximation to the multinomial case that

\[
P[\bar{X} \geq 0; \bar{Y} \geq 0] \sim \rho \quad .
\]

The existence and uniqueness of \(T_i(\lambda)\) and \(\sigma_i(\lambda)\) \(i = 1,2\) are guaranteed
analogously to Section D.1 of Chapter III by the assumptions.

Let \(C(s)\) be a region composed of mutually exclusive squares \(A_i(s) = \{(x,y) : a_i(s) < x \leq a_i(s) + s, b_i(s) < y \leq b_i(s) + s\}\) such that \(P[A_i(s)] > 0\)
for each \(i\) and the set \(\bigcup_i A_i(s) \setminus \Omega\) has probability zero. Let \((X(s),Y(s))\) have 
the probability distribution that assigns probability \(P[A_i(s)] = \Pi_i(s) > 0\)
to \((a_i(s),b_i(s))\), i.e., the probability of the square is assigned to the 
lower left hand corner. Let \(m(s)(t_1, t_2)\) be the moment generating function 
of \((X(s),Y(s))\) and let \(A(s)\) and \(B(s)\) be defined analogously to the definition 
of \(A\) and \(B\) in Section D.1 of Chapter III. Finally, let \(\rho(s)\) be defined
analagously to the definition of \( p \) in Section D.1 of Chapter III.

Then applying arguments essentially used by Chernoff (3) in the one-dimensional case, it is clear that, given \( \epsilon > 0 \), there exists an \( s \) such that

\[
|p - p(s)| < \epsilon \quad (11)
\]

Similarly, let \((X(r), Y(r))\) have the probability distribution \( P(r) \) that assigns probability \( P(A_i(r)) \equiv \prod_{i} (r) > 0 \) to \((a_i(r) + r, b_i(r) + r)\), i.e.,

the probability of each square is assigned to the upper right hand corner.

Let \( m(r)(t_1, t_2) \) be the moment generating function of \((X(r), Y(r))\); define as well the corresponding \( A(r), B(r) \) and \( p(r) \).

Then given \( \epsilon > 0 \) there exists an \( r \) such that

\[
|p - p(r)| < \epsilon \quad (12)
\]

Now suppose we take a sample of size \( N \); define \( \bar{X}_N = \sum_i a_i(s) n_i \) and \( \bar{Y}_N = \sum_i b_i(s) n_i \), \( n_i \) being the number of observations that fall into the square \( A_i(s) \). Probabilities of various realizations of the \( n_i \) 's are computed according to the distribution \( p(s) \). Similarly define \( \bar{X}_N = \sum_i a_i(r) m_i \) and \( \bar{Y}_N = \sum_i b_i(r) m_i \), \( m_i \) being the number of observations that fall into \( A_i(r) \) with the distribution in question being \( P(r) \). Then:

\[
p(s) \sim P\{\sum_i a_i(s) n_i \geq 0; \ \sum_i b_i(s) n_i \geq 0\}
= P\{\bar{X}_N \geq 0; \ \bar{Y}_N \geq 0\}
\leq P\{\bar{X} \geq 0; \ \bar{Y} \geq 0\}
\leq P\{\bar{X}_N \geq 0; \ \bar{Y}_N \geq 0\}
\]
the first and last equivalences following from Section D.1 of Chapter III. 
Now equation (10) follows from equations (11), (12) and (13).

E. Multivariate Large Deviations for Exponential Families

Consider an exponential family with densities

\[ p_\theta(x) = \exp\{T(x)'\theta - \psi(\theta)\} h(x) \] \quad \theta \in \Theta, \ T(x) \in X

where \( \theta \) and \( T(x) \) are \( k \)-dimensional vectors \( \theta = (\theta_1, \ldots, \theta_k) \) and \( T(x) = (T_1(x), \ldots, T_k(x)) \), \( \Theta \) is the natural parameter space and \( X \) is a set in \( E_k \) which is not of lower dimension (i.e., is not a hyperplane). \( \Theta \) is assumed open.

Let \( m_\theta(t) \) be the natural logarithm of the moment generating function of \( T(x) \) for \( \theta \in \Theta \). Then we have

Lemma 6: \( m_\theta(t) = \psi(\theta + t) - \psi(\theta) \) \quad \text{if} \quad \theta \in \Theta.

Pf.: The moment generating function of \( T(x) \) is:

\[
\exp\left\{ \sum_{j=1}^{k} T_j(x)(t_j + \theta_j) - \psi(\theta) \right\} h(x) dx = \exp\left\{ \psi(t + \theta) - \psi(\theta) \right\} h(x) dx
\]

\[
= \exp\left\{ \psi(t + \theta) - \psi(\theta) \right\} \cdot h(x) dx
\]

Consider, analogously to Section D, the map

\[ a = \text{gradient (} m_\theta(t) \text{)} \]

i.e.,

\[ a = \text{gradient } [\psi(\theta + t) - \psi(\theta)] \]

\[ a \mid_{\Theta} = \text{gradient } \psi(\theta + t) \] \quad \text{if} \quad \theta \in \Theta - \Theta
\[ \frac{\partial}{\partial \theta} f(\theta) \mid_{\theta_0} = \text{gradient } \psi(\theta) \mid_{\theta_0} = f(\theta) \quad \theta \in \Theta \]

where
\[ \Theta = \{ \theta : \theta + \theta_0 \in \Theta \} \quad \text{and} \quad \theta = I + \theta_0 \quad . \]

It follows from Theorem 9 (9, p.52) that second order derivatives of \( m_\theta(t) \) may be obtained by differentiating under the integral sign. Then analogously to Section D of Chapter III the function \( f \) maps \( \theta \in \Theta \) onto its range, \( A \), in a one to one manner. Thus \( f^{-1} \) is well-defined over \( A \).

It follows from the general considerations of Borovkov and Rogozin that, as in the cases treated in Section D of Chapter III, the large deviations of \( \tilde{T} = (\tilde{T}_1, \ldots, \tilde{T}_k)(\tilde{T}_j = \frac{1}{n} \sum_{i=1}^{n} T_j(x_i)) \) to a point \( \theta \) for parameter \( \theta_0 \), are governed by
\[ I(\theta, \theta) = \inf_{\tilde{T}} \left[ \theta' \tilde{T} - m_\theta(\tilde{T}) \right] \quad . \]

It will be of use in Chapter IV to note the explicit expression
\[ I(\theta, \theta) = \theta' [f^{-1}(\theta) - \theta] + \psi(\theta) - \psi(f^{-1}(\theta)) \geq 0 \]

which is derived below.

The minimizing \( \tilde{T}_0 \) is the solution of
\[ \theta = \text{gradient } m_\theta(t) \mid_{\tilde{T}_0} \quad , \]
i.e.,
\[ \tilde{T}_0 = f^{-1}(\theta) - \theta \quad . \]

Hence
\[ I(\theta, \theta) = \theta' [f^{-1}(\theta) - \theta] - \left[ \psi(f^{-1}(\theta) - \theta + \theta) - \psi(\theta) \right] \]
Finally, we note that, from the discussion of existence in Section D of Chapter III, $A = C(X)$. 

\[ a'[f^{-1}(a) - a] + \psi(a) - \psi(f^{-1}(a)) \]
IV. ASYMPTOTIC OPTIMALITY AND MINIMAXNESS

A. Asymptotic Optimality of a Certain Likelihood Ratio Procedure

Of the two difficulties mentioned in Section B of Chapter 11, only the second has been removed by the procedure suggested there. In order to study the asymptotic properties of the \( \delta_0 \) obtained in Section B of Chapter 11, and avoid the first difficulty we shall introduce a new test function \( \tilde{\delta}(x) \) which is suggested by the three possible \( \delta_0 \)'s. We would like to show that, asymptotically at least its risk is equivalent to that of \( \delta_0 \) in the sense that:

\[
\max R_1(\delta,j) \leq \inf \max R_1(\delta,j) \leq \max R_1(\delta,j)
\]

where

\[
\tilde{\delta}(x_1,\ldots,x_n) = \begin{cases} \max_n \frac{\prod f_1(x_i) \prod f_2(x_i)}{\prod f_0(x_i)} & \text{if } \frac{\prod f_1(x_i) \prod f_2(x_i)}{\prod f_0(x_i)} > 1 \\ 1 & \text{otherwise} \end{cases}
\]

We will actually show that:

\[
\lim_{n \to \infty} \frac{1}{n} \max \left( \min_{\delta \in B'} \max_{j} R_1(\delta,j) \right) = \frac{1}{n} \max R_1(\delta,j) = 1
\]

But

\[
\max_{j} \inf_{\delta} R_1(\delta,j) \leq \max \inf_{\delta} R_1(\delta,p) = \inf_{p} \max R_1(\delta,p) = \inf_{\delta} \max R_1(\delta,j).
\]

Thus if (3) holds

\[
\lim_{n \to \infty} \frac{1}{n} \max \min_{\delta \in B} R_1(\delta,j) \leq 1
\]
On the other hand since $\delta \in B'$

$$\max R_1(\delta, j) \geq \inf \max R_1(\delta, j)_{\delta \in B'}$$

or

$$\lim_{n \to \infty} \inf_{\delta \in B'} \max R_1(\delta, j) \geq 1$$

and (5) and (6) are equivalent to (1). What we have shown is that if

$$\max R_1(\delta, j) \leq 1$$

and

then (1) holds. If equation (1) holds we shall say that $\delta$ is asymptotically minimax. If equation (7) holds (and this is what we will prove under certain conditions) we shall say that $\delta$ is asymptotically optimal because it is asymptotically equivalent in terms of the risk to one of the Neyman-Pearson test functions. Thus, if $\delta$ is asymptotically optimal then it's asymptotically minimax.

To evaluate the risk we first note that $R_1(\delta, j)$ is the sum of the probabilities of Type I and II errors in testing $H_0: f_0$ vs. $H_A: f_{j+1}$. Thus letting $L(f_j) = \prod_{i=1}^{n} f_j(x_i)$, we may rewrite $R_1(\delta, j)$ as a sum of bivariate probabilities:

$$R_1(\delta, j) = P_{f_{j+1}} \left\{ \max \left\{ \frac{L(f_1), L(f_2)}{L(f_0)} \right\} < 1 \right\} + P_{f_0} \left\{ \max \left\{ \frac{L(f_1), L(f_2)}{L(f_0)} \right\} > 1 \right\}$$

$$= P_{f_{j+1}} \left\{ \frac{L(f_1)}{L(f_2)} > 1; \frac{L(f_1)}{L(f_0)} < 1 \right\} + P_{f_{j+1}} \left\{ \frac{L(f_1)}{L(f_2)} < 1; \frac{L(f_2)}{L(f_0)} < 1 \right\}$$
The formula we are using here is:

\[
P\left[ \frac{\max(A, B)}{C} > 1 \right] = P\left[ \max(A, B) = A \text{ and } \frac{A}{C} > 1 \right] + P\left[ \max(A, B) = B \text{ and } \frac{B}{C} > 1 \right].
\]

Similarly we may express the risks of the Neyman-Pearson tests as:

\[
R_{1}(j, \delta_{NP}, j+1, j) = P_{f} \left\{ \frac{L(f_{1})}{L(f_{2})} > 1; \frac{L(f_{1+1})}{L(f_{0})} < 1 \right\} + P_{f} \left\{ \frac{L(f_{1})}{L(f_{2})} < 1; \frac{L(f_{1+1})}{L(f_{0})} > 1 \right\}
\]

\[
+ P_{f} \left\{ \frac{L(f_{1})}{L(f_{2})} > 1; \frac{L(f_{1+1})}{L(f_{0})} > 1 \right\} + P_{f} \left\{ \frac{L(f_{1})}{L(f_{2})} < 1; \frac{L(f_{1+1})}{L(f_{0})} < 1 \right\}
\]

\[j = 0,1 \quad (9)\]

We note that:

1) each of these bivariate probabilities could, by reversing likelihoods, be stated as the probability that two likelihood ratios are greater than 1;
2) by taking logs each term could be converted into the probability that two sums of independent, identically distributed random variables (or, dividing by n, two means) are greater than zero;
3) in equation (8) (with j=0 or 1) and equation (9) (with j= 0 or 1) two terms are identical.

We shall say that one bivariate probability dominates another if the p for the former is greater than the p for the latter. Then its clear that:
Lemma 1: If identical terms dominate \( \max_j R_1(\delta_j) \) and \( \max_j R_1(\delta_{NP,j+1}) \) then as \( n \to \infty \)

\[
\frac{\max_j R_1(\delta_j)}{\max_j R_1(\delta_{NP,j+1})} \to 1.
\]

The above lemma, while useful, is rather restrictive. A more general case is that in which the dominant terms in \( \max_j R_1(\delta_j) \) and \( \max_j R_1(\delta_{NP,j+1}) \) have the same \( p \) but are not identical. In this situation

\[
\lim_{n \to \infty} \frac{\max_j R_1(\delta_j)}{\max_j R_1(\delta_{NP,j+1})} = \lim_{n \to \infty} \frac{\rho^n}{n}
\]

which generally will not be 1.

This could be avoided if we had, instead of using \( R_1(\delta_j) \) in Section B of Chapter II, used

\[ R_3(\delta_j) = \ln R_1(\delta_j) \]  \hspace{1cm} (10)

as our risk function. Using only the fact that \( \ln x \) is a monotonic increasing function we could now, by mimicking the above methods, prove everything that we did there with this new risk function. In addition we can show that for any \( \rho, 0 < \rho < 1 \)

\[
\lim_{n \to \infty} \frac{\ln[\rho^n]}{n \ln[\rho^n]} = \lim_{n \to \infty} \frac{n \ln \rho + \ln 1}{n \ln \rho + \ln 1} = 1 \hspace{1cm} (11)
\]
First of all the limit must exist since by Lemmas 4 and 5 (Section C of Chapter III) \( p^n \) is the dominant part of \( P[\hat{x}_n \geq 0; \hat{y}_n \geq 0] \), i.e., the dominant term of the numerator which is a sum of two terms is the same as the dominant term of the denominator. Secondly if

\[
\lim_{n \to \infty} \frac{n \ln p + t_n}{n \ln p + t_1} = c > 1
\]

then for large \( n \)

\[
n \ln p + t_n \approx c [n \ln p + t_1]
\]

or

\[
p^n \approx p^n \left( \frac{t_1}{t_n} \right)^c
\]

which since \( C \) is a constant contradicts the fact that \( n \ln p \) is the dominating term of both the numerator and denominator of the above equation in which the risk \( R_3 \) was used. Thus (II) holds, i.e., (I) holds more generally with the risk being \( R_3 \) instead of \( R_1 \).

However in the two examples we present we will be able to show that \( \tilde{\delta} \) is asymptotically optimal by using either \( R_1 \) or the weaker \( R_3 \).

A particularly simple large deviation formula obtains if we consider the bivariate normal distribution, say of a random variable \((X_1, X_2)\) with mean \((0,0)\) and covariance matrix \(I\). If we take a sample of size \( n \), \( \sqrt{n}(\bar{x}, \bar{y}) \) has the same distribution. We may compute the probability of any "wedge" in \( E_2 \) produced by the intersection of two half spaces by finding the distance, say \( D \), from the origin to the point of the wedge closest to the origin, i.e.

**Theorem 1:** Suppose \( \sqrt{n}(\bar{x}_1, \bar{x}_2) \) has a bivariate normal distribution with mean 0 and covariance matrix 1. The probability of any wedge not containing the
origin can be approximated by

\[ \beta(n)e^{-\frac{n^2}{2}} \quad (12) \]

where for large enough \( n \) there exist constants \( K_1, K_2 \) and \( \varepsilon > 0 \) such that

\[ \frac{K_1}{n^{1+\varepsilon}} < \beta(n) < \frac{K_2}{n^{1/2-\varepsilon}} \]

**Pf.** If the point closest to \((0,0)\) say \((D_1,D_2)\) is on either the \(\sqrt{n}x_1\) or \(\sqrt{n}x_2\) axis we can rotate the wedge by \(45^\circ\) (i.e., multiply \(\sqrt{n}(\bar{x}_1,\bar{x}_2)\) by \(\frac{1}{\sqrt{2}}\begin{pmatrix} 1 & 1 \\ -1 & 1 \end{pmatrix}\) which will not change the distribution or the probability). Thus we can assume \(D_i \neq 0\) \(i=1,2\). There are two possibilities. (1) \((D_1,D_2)\) is on an edge of the wedge, or (2) \((D_1,D_2)\) is at the vertex of the wedge.

(1) Suppose the equations of the two half spaces are

\[ a_1\sqrt{n} \bar{x}_1 + a_2\sqrt{n} \bar{x}_2 \geq a_3 \]

\[ a_4\sqrt{n} \bar{x}_1 + a_5\sqrt{n} \bar{x}_2 \leq a_6 \]

We can assume without loss of generality that \((D_1,D_2)\) is on the edge

\[ a_1\sqrt{n} \bar{x}_1 + a_2\sqrt{n} \bar{x}_2 = a_3, \quad \text{i.e., } a_1D_1 + a_2D_2 = a_3 \]

From above we see that \(D' = D_1D_2\) and \(D = \sqrt{D_1^2 + D_2^2}\) will characteristically be constant multiples of \(\sqrt{n}\) (e.g., \(\bar{x}_1 + \bar{x}_2 = 1 \Leftrightarrow \sqrt{n} \bar{x}_1 + \sqrt{n} \bar{x}_2 = \sqrt{n}\)). Clearly

\[ P[a_1\sqrt{n} \bar{x}_1 + a_2\sqrt{n} \bar{x}_2 \geq a_3; a_4\sqrt{n} \bar{x}_1 + a_5\sqrt{n} \bar{x}_2 \leq a_6] \leq P[a_1\sqrt{n} \bar{x}_1 + a_2\sqrt{n} \bar{x}_2 \geq a_3] \]

Now we perform a rotation \((\sqrt{n} \bar{x}_1, \sqrt{n} \bar{x}_2) \rightarrow (y_1,y_2)\) so that the half plane whose probability we're calculating becomes \([(y_1,y_2): y_1 \geq c]\). We know that the probability of the half plane and the (bivariate normal) distribution are unchanged by any rotation. Also, since \((0,0)\) was not in the wedge and
since \((D_1, D_2)\) is on an edge, it's clear (geometrically) that \((0,0)\) is not in the half-plane \(a_1\sqrt{n} \bar{x}_1 + a_2\sqrt{n} \bar{x}_2 \geq a_3\) which means \(c > 0\). Furthermore since distances remain invariant under rotations \(c = D\), the distance from the origin to the half-plane. This means that:

\[
P\{a_1\sqrt{n} \bar{x}_1 + a_2\sqrt{n} \bar{x}_2 \geq a_3\} = P\{y_1 \geq D\}.
\]

Since \((y_1, y_2)\) has the bivariate normal distribution with mean \(0\) and covariance matrix \(I\) \(y_1\) has the standard \(N(0,1)\) distribution and if we use a tail area approximation for this distribution (6, p.166)

\[
P\{y_1 \geq D\} \approx \frac{1}{\sqrt{2\pi D}} e^{-D^2/2}
\]

i.e.,

\[
P\{a_1\sqrt{n} \bar{x}_1 + a_2\sqrt{n} \bar{x}_2 \geq a_3:\ a_4\sqrt{n} \bar{x}_1 + a_5\sqrt{n} \bar{x}_2 \leq a_6\} \approx \frac{1}{\sqrt{2\pi D}} e^{-D^2/2}.
\]  

(13)

On the other hand, since \((D_1, D_2)\) is on an edge of the wedge we can imbed a small rectangle with vertices \((D_1, D_2), (D_1 + \Delta_1, D_2), (D_1, D_2 + \Delta_2), (D_1 + \Delta_1, D_2 + \Delta_2)\) in the wedge. Since \(D_1 = \sqrt{n} \cdot \text{constant} \) we take \(\Delta_1 = \sqrt{n} \cdot \text{constant}\). We know that the probability of the above rectangle is smaller than the probability of the wedge. Since \(\sqrt{n} \bar{x}_1\) and \(\sqrt{n} \bar{x}_2\) are independent variables the probability of this rectangle is equal to the product of the (univariate) probabilities of its sides. Suppose \(D_1 > 0\). Then using the above approximation twice:

\[
P[D_1 < \sqrt{n} \bar{x}_1 < D_1 + \Delta_1] = P[D_1 < \sqrt{n} \bar{x}_1] - P[D_1 + \Delta_1 < \sqrt{n} \bar{x}_1]
\]

\[
= \frac{1}{\sqrt{2\pi D_1}} e^{-D_1^2/2} - \frac{1}{\sqrt{2\pi (D_1 + \Delta_1)^2}} e^{-(D_1 + \Delta_1)^2/2}
\]
55

\[
\begin{align*}
    &\frac{1}{\sqrt{2\pi D_1}} e^{-D_1^2/2} [1 - \frac{D_1}{D_1 + \Delta_1} e^{1/2 \left[D_1^2 - (D_1 + \Delta_1)^2\right]}] \\
    &\geq \frac{1}{\sqrt{2\pi D_1}} e^{-D_1^2/2}
\end{align*}
\]

because \( \Delta_1 = \sqrt{n} \delta \delta > 0 \), which makes the second term on the right hand side negligible with respect to the first when \( n \) is large. If \( D_1 < 0 \)

\[
P[D_1 < \sqrt{n} \bar{X}_1 < D_1 + \Delta_1] \geq P[-D_1 - \Delta_1 < \sqrt{n} \bar{X}_1 < -D_1]
\]

\[
= \frac{1}{\sqrt{2\pi(-D_1-\Delta_1)}} e^{-(-D_1-\Delta_1)^2/2}
\]

\[
\geq \frac{1}{\sqrt{2\pi |D_1|}} e^{-D_1^2/2}
\]

Similarly if \( D_2 > 0 \)

\[
P[D_2 < \sqrt{n} \bar{X}_2 < D_2 + \Delta_2] \geq \frac{1}{\sqrt{2\pi D_2}} e^{-D_2^2/2}
\]

Therefore

\[
P[a_1 \sqrt{n} \bar{X}_1 + a_2 \sqrt{n} \bar{X}_2 < a_2; a_4 \sqrt{n} \bar{X}_1 + a_5 \sqrt{n} \bar{X}_2 < a_6] \geq \frac{1}{2\pi |D_1 D_2|} e^{-\left(D_1^2 + D_2^2\right)/2}
\]

\[
= \frac{1}{2\pi |D_1 D_2|} e^{-D^2/2} \quad (14)
\]

Equations (13) and (14) imply

\[
\frac{1}{2\pi |D_1 D_2|} e^{-D^2/2} \leq P[a_1 \sqrt{n} \bar{X}_1 + a_2 \sqrt{n} \bar{X}_2 \geq a_3; a_4 \sqrt{n} \bar{X}_1 + a_5 \sqrt{n} \bar{X}_2 \leq a_6]
\]

\[
\leq \frac{1}{\sqrt{2\pi D}} e^{-D^2/2}
\]
Since we can set $D_i = \sqrt{n} L_i$, $i = 1, 2$, $L_i$ = constant, we may write:

$$P\{a_1\sqrt{n} \bar{x}_1 + a_2\sqrt{n} \bar{x}_2 \geq a_3; a_4\sqrt{n} \bar{x}_1 + a_5\sqrt{n} \bar{x}_2 \leq a_6\} = \beta(n) e^{-\frac{n(L_1^2 + L_2^2)}{2}}$$

where $\frac{1}{2\pi n L_1 L_2} \leq \beta(n) \leq \frac{1}{\sqrt{2\pi(L_1^2 + L_2^2)n}}$.

(2) Now the closest point of the wedge is at the vertex. Since the origin is not in the wedge it is not in at least one of the half-planes whose intersection forms the wedge. Proceeding as in (1) we may use the probability of this half-plane as an upper bound for the probability of the wedge (equation (13)).

On the other hand by taking points which are inside the wedge and just a little farther from the origin than the closest point we can construct a sequence of rectangles of the type considered in (1) such that the approximate probability of one of them is

$$\frac{1}{2\pi |D_1 D_2|} e^{-\frac{D_i^2}{2}}$$

where if $D_1 > 0$, $D_2 > 0$, $D_1 D_2 \leq D_1 D_2$, and $D_k \perp D$ as $k \to \infty$ or if $D_1 < 0$, $D_2 > 0$, $D_1 \perp D_2$, $D_2 \perp D_2$, and $D_k \perp D$ as $k \to \infty$ etc. Therefore for any $D_i > D$ and corresponding $D_i \perp (D_i, D_2)$ is in the wedge

$$P\{a_1\sqrt{n} \bar{x}_1 + a_2\sqrt{n} \bar{x}_2 \geq a_3; a_4\sqrt{n} \bar{x}_1 + a_5\sqrt{n} \bar{x}_2 \leq a_6\} \geq \frac{1}{2\pi |D_1 D_2|} e^{-(D_i)^2/2}$$

which means that equation (12) is a valid approximation in this case also.

q.e.d.

If we wished to calculate these probabilities for a bivariate normal distribution with non-zero mean $\mu$, we would just translate the $(\sqrt{n} \bar{x}_1, \sqrt{n} \bar{x}_2)$
plane so that the origin was at \((\sqrt{n} \mu_1, \sqrt{n} \mu_2)\) to reduce this to the previous case. This corresponds to measuring distances (i.e., \(D_1, D_2, D\)) from \(\sqrt{n} \mu\) in the original \((\sqrt{n} \bar{x}_1, \sqrt{n} \bar{x}_2)\) space.

Utilizing this result we can now prove equation (3) for a special case.

**Theorem 2:** Suppose \(f_i(x_1, x_2)\) \(i = 0, 1, 2\) are bivariate normal distributions with distinct means \(\mu_i\) \(i = 0, 1, 2\) and covariance matrix \(I\). Also suppose that \(d(\mu_0, \mu_1) \neq d(\mu_0, \mu_2)\) where \(d\) represents Euclidean distance. Then

\[
\max_j R_1(\delta, j) \frac{1}{\max_j R_1(\delta_{NP, j} + 1, j)} \rightarrow 1 \text{ as } n \rightarrow \infty .
\]  

**Pf.:** We may without loss of generality suppose that \(d(\mu_0, \mu_1) < d(\mu_0, \mu_2)\)

From the above we see that for a sample of size \(n\) we must consider the eight sets in the sample space satisfying the following pairs of inequalities

\[
\begin{align*}
\frac{L(f_1)}{L(f_2)} > 1; & \frac{L(f_1)}{L(f_0)} < 1 \\
\frac{L(f_1)}{L(f_2)} < 1; & \frac{L(f_1)}{L(f_0)} < 1 \\
\frac{L(f_1)}{L(f_2)} > 1; & \frac{L(f_1)}{L(f_0)} > 1 \\
\frac{L(f_1)}{L(f_2)} < 1; & \frac{L(f_1)}{L(f_0)} > 1 \\
\frac{L(f_1)}{L(f_2)} > 1; & \frac{L(f_1)}{L(f_0)} < 1 \\
\frac{L(f_1)}{L(f_2)} < 1; & \frac{L(f_1)}{L(f_0)} < 1 \\
\frac{L(f_1)}{L(f_2)} > 1; & \frac{L(f_1)}{L(f_0)} > 1 \\
\end{align*}
\]
\[ \frac{L(f_1)}{L(f_2)} < \frac{L(f_2)}{L(f_0)} > 1. \]

We note that
\[ \frac{L(f_1)}{L(f_2)} > 1 \iff \exp\left[-\frac{1}{2} \sum_{i=1}^{n} (X_i - \mu_1)'(X_i - \mu_2) + \frac{1}{2} \sum_{i=1}^{n} (X_i - \mu_2)'(X_i - \mu_1) \right] > 1 \]
\[ \iff \sum_{i=1}^{n} (X_i - \mu_2)'(X_i - \mu_2) > \sum_{i=1}^{n} (X_i - \mu_1)'(X_i - \mu_1) \]
\[ \iff d(\bar{x}_2, \bar{\mu}_2) > d(\bar{x}_1, \bar{\mu}_1) \]
\[ \iff d(\sqrt{n} \bar{x}_2, \sqrt{n} \bar{\mu}_2) > d(\sqrt{n} \bar{x}_1, \sqrt{n} \bar{\mu}_1) \]

Denoting the \(i\)th set above \(A_1\) we may write:

\[ A_1 = \{ \bar{x} : d(\sqrt{n} \bar{x}, \sqrt{n} \bar{\mu}_1) < d(\sqrt{n} \bar{x}, \sqrt{n} \bar{\mu}_2) ; d(\sqrt{n} \bar{x}, \sqrt{n} \bar{\mu}_1) > d(\sqrt{n} \bar{x}, \sqrt{n} \bar{\mu}_0) \} \]
\[ A_2 = \{ \bar{x} : d(\sqrt{n} \bar{x}, \sqrt{n} \bar{\mu}_1) > d(\sqrt{n} \bar{x}, \sqrt{n} \bar{\mu}_2) ; d(\sqrt{n} \bar{x}, \sqrt{n} \bar{\mu}_1) > d(\sqrt{n} \bar{x}, \sqrt{n} \bar{\mu}_0) \} \]
\[ A_3 = \{ \bar{x} : d(\sqrt{n} \bar{x}, \sqrt{n} \bar{\mu}_1) < d(\sqrt{n} \bar{x}, \sqrt{n} \bar{\mu}_2) ; d(\sqrt{n} \bar{x}, \sqrt{n} \bar{\mu}_1) < d(\sqrt{n} \bar{x}, \sqrt{n} \bar{\mu}_0) \} \]
\[ A_4 = \{ \bar{x} : d(\sqrt{n} \bar{x}, \sqrt{n} \bar{\mu}_1) > d(\sqrt{n} \bar{x}, \sqrt{n} \bar{\mu}_2) ; d(\sqrt{n} \bar{x}, \sqrt{n} \bar{\mu}_1) < d(\sqrt{n} \bar{x}, \sqrt{n} \bar{\mu}_0) \} \]
\[ A_5 = \{ \bar{x} : d(\sqrt{n} \bar{x}, \sqrt{n} \bar{\mu}_1) < d(\sqrt{n} \bar{x}, \sqrt{n} \bar{\mu}_2) ; d(\sqrt{n} \bar{x}, \sqrt{n} \bar{\mu}_2) > d(\sqrt{n} \bar{x}, \sqrt{n} \bar{\mu}_0) \} \]
\[ A_6 = \{ \bar{x} : d(\sqrt{n} \bar{x}, \sqrt{n} \bar{\mu}_1) > d(\sqrt{n} \bar{x}, \sqrt{n} \bar{\mu}_2) ; d(\sqrt{n} \bar{x}, \sqrt{n} \bar{\mu}_2) > d(\sqrt{n} \bar{x}, \sqrt{n} \bar{\mu}_0) \} \]
\[ A_7 = \{ \bar{x} : d(\sqrt{n} \bar{x}, \sqrt{n} \bar{\mu}_1) < d(\sqrt{n} \bar{x}, \sqrt{n} \bar{\mu}_2) ; d(\sqrt{n} \bar{x}, \sqrt{n} \bar{\mu}_2) < d(\sqrt{n} \bar{x}, \sqrt{n} \bar{\mu}_0) \} \]
\[ A_8 = \{ \bar{x} : d(\sqrt{n} \bar{x}, \sqrt{n} \bar{\mu}_1) > d(\sqrt{n} \bar{x}, \sqrt{n} \bar{\mu}_2) ; d(\sqrt{n} \bar{x}, \sqrt{n} \bar{\mu}_2) < d(\sqrt{n} \bar{x}, \sqrt{n} \bar{\mu}_0) \} \]

With this notation:
\[ R_1(\delta, 0) = P_{f_1}[A_1] + P_{f_1}[A_2] + P_{f_0}[A_3] + P_{f_0}[A_8] \]

Now since \(d^2(\sqrt{n} x, \sqrt{n} y) = nd^2(x, y)\) and since the minimum distance from \(\sqrt{n} \mu_1\) to \(A_1\) is \(\frac{1}{2}d(\sqrt{n} \mu_1, \sqrt{n} \mu_0) = \frac{\sqrt{n}}{2}d(\mu_1, \mu_0)\)
This minimum distance is achieved at one point, \( b \), halfway along the line from \( \sqrt{n} \mu_0 \) to \( \sqrt{n} \mu_1 \). Consider the set \( A_6 \). \( \tilde{x} \in A_6 \Rightarrow d(\sqrt{n} \tilde{x}, \sqrt{n} \mu_1) > d(\sqrt{n} \tilde{x}, \sqrt{n} \mu_2) \). Since \( d(\mu_0, \mu_2) > d(\mu_0, \mu_1) \), \( \sqrt{n} \mu_2 \) is outside the circle of radius \( \sqrt{n} d(\mu_0, \mu_1) \) centered at \( \sqrt{n} \mu_0 \). On the other hand \( b \) is inside this circle and \( d(b, \sqrt{n} \mu_0) = \sqrt{n} d(b, \mu_0) \). Thus \( d(b, \sqrt{n} \mu_1) > d(b, \sqrt{n} \mu_2) \), i.e., \( b \notin A_6 \). Also \( A_6 \subset A_9 = \{ \tilde{x} : d(\sqrt{n} \tilde{x}, \sqrt{n} \mu_1) > d(\sqrt{n} \tilde{x}, \sqrt{n} \mu_2) \} \) and the minimum distance from \( \sqrt{n} \mu_1 \) to \( A_9 \) is \( \sqrt{n} d(\mu_0, \mu_1) \) which again is achieved only at \( b \). Thus the minimum distance from \( \sqrt{n} \mu_1 \) to \( A_6 \) is greater than \( \sqrt{n} d(\mu_0, \mu_1) \):

\[
P_{f_1}(A_6) \sim K_4 > \frac{1}{2} d(\mu_0, \mu_1)
\]

By similar reasoning:

\[
P_{f_0}(A_3) \sim \frac{1}{2} d(\mu_0, \mu_1)
\]

From the second condition defining \( A_8 \):

\[
P_{f_0}(A_3) \sim K_5 \geq \frac{1}{2} d(\mu_2, \mu_0)
\]

Similarly:

\[
R_1(\delta, 1) = P_{f_2}(A_1) + P_{f_2}(A_6) + P_{f_0}(A_3) + P_{f_0}(A_8)
\]

and from the second condition defining \( A_1 \):

\[
P_{f_2}(A_1) \sim K_6 \geq \frac{1}{2} d(\mu_0, \mu_2)
\]

Also from the second condition defining \( A_6 \):
There are two dominant terms in $R(\delta, 0)$ each of magnitude $\frac{1}{2}d(\mu_0, \mu_1)$ whereas there is only one such term in $R(\delta, 1)$ which means that for large $n$

$$\max_j R(\delta, j) = R(\delta, 0).$$

Turning to the Neyman-Pearson risks

$$R_1(\delta_{NP, 1}, 0) = P_{f_1}[A_1] + P_{f_1}[A_2] + P_{f_0}[A_3] + P_{f_0}[A_4].$$

If we proceed similarly to when we evaluated $P_{f_1}[A_6]$ we obtain:

$$P_{f_1}[A_2], K_8 > \frac{1}{2}d(\mu_0, \mu_1),$$

$$P_{f_0}[A_4], K_9 > \frac{1}{2}d(\mu_0, \mu_1).$$

Similarly

$$R_1(\delta_{NP, 2}, 1) = P_{f_2}[A_5] + P_{f_2}[A_6] + P_{f_0}[A_7] + P_{f_0}[A_8].$$

From the second condition defining $A_5$:

$$P_{f_2}[A_5], K_{10} > \frac{1}{2}d(\mu_0, \mu_2).$$

From the second condition defining $A_7$:

$$P_{f_0}[A_7], K_{11} > \frac{1}{2}d(\mu_0, \mu_2).$$

Thus there are two dominating terms in $R_1(\delta_{NP, 1}, 0)$ each of magnitude $\frac{1}{2}d(\mu_0, \mu_1).$ All terms in $R_1(\delta_{NP, 2}, 1)$ are of smaller magnitude, i.e., for large $n$

$$\max_j R_1(\delta_{NP, j+1}, j) \leq R(\delta_{NP, 1}, 0) \leq P_{f_1}[A_1] + P_{f_0}[A_3].$$
In addition when \( n \) is large

\[
\max_{j} R_{1}(\delta_{j}) = R(\delta_{0}) \\
\approx p_{f_{1}}[A_{1}] + p_{f_{0}}[A_{3}]
\]

The last two equations imply:

\[
\lim_{n \to \infty} \frac{\max_{j} R_{1}(\delta_{j})}{\max_{j} R_{1}(\delta_{NP_{j} + 1})} = 1
\]

If \( d(\mu_{0,1}, \mu_{2}) < d(\mu_{0,1}, \mu_{1}) \) we use the same procedure to obtain this result.

q.e.d.

It is natural to assume that this result also holds when \( d(\mu_{0,1}, \mu_{1}) =

Thus \( R(\delta_{0}) \) and \( R(\delta_{1}) \) each have three dominant terms of the same magnitude. Furthermore two of these terms are identical and the third terms \( (p_{f_{2}}[A_{6}]) \) and \( p_{f_{1}}[A_{3}] \) are equal by the symmetry of their definitions and of the problem, i.e., \( d(\mu_{0,1}, \mu_{1}) = d(\mu_{0,1}, \mu_{2}) \). Thus:
\[ \max_j R_1(\bar{s},j) = R_1(\bar{s},0) \]
\[ = R_1(\bar{s},1) \]

However, if we calculate the probabilities involved in \( R_1(\delta_{NP,j+1,j}) \) we get:

\[ P_{f_1}[A_2] \sim K_{14} > \frac{1}{2} d(\mu_0, \mu_1) \]
\[ P_{f_0}[A_4] \sim K_{15} > \frac{1}{2} d(\mu_0, \mu_2) \]
\[ P_{f_2}[A_5] \sim K_{16} > \frac{1}{2} d(\mu_0, \mu_2) \]
\[ P_{f_0}[A_7] \sim K_{17} > \frac{1}{2} d(\mu_0, \mu_2) \]

\( R_1(\delta_{NP,1,0}) \) and \( R_1(\delta_{NP,2,1}) \) each have two dominant terms of the same order. Furthermore by using the symmetry argument above we can show that these probabilities are equal pairwise (i.e. \( P_{f_0}[A_3] = P_{f_0}[A_8] \) and \( P_{f_1}[A_1] = P_{f_2}[A_6] \)) and we have:

\[ \max_j R_1(\delta_{NP,j+1,j}) = R_1(\delta_{NP,1,0}) \]
\[ = R_1(\delta_{NP,2,1}) \]

If we try to verify equation (5) in this case

\[ \lim_{n} \frac{\max_j R_1(\delta_{NP,j+1,j})}{n} \]
\[ = \lim_{n} \frac{R_1(\delta,0)}{n} \]
\[ = \lim_{n} \frac{P_{f_1}[A_1] + P_{f_0}[A_3] + P_{f_0}[A_8]}{P_{f_1}[A_1] + P_{f_0}[A_3]} \]
\[ \lim_{n \to \infty} \frac{P_{f_1}(A_1) + 2P_{f_0}(A_3)}{P_{f_1}(A_1) + P_{f_0}(A_3)} \neq 1 \]

since both of these probabilities are of the same order. We note that this limit will be 1 if we use the risk function \( R_3(\delta, j) \) instead of \( R_1(\delta, j) \).

We will illustrate Theorem 2 with an example. Let \( f_0(x) = f_0(x_1, x_2) \) be a \( N[\begin{pmatrix} 1 \\ 1 \end{pmatrix}, \Sigma] \) density, \( f_1(x) \) a \( N[\begin{pmatrix} 1 \\ 1 \end{pmatrix}, \Sigma] \) density and \( f_2(x) \) a \( N[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \Sigma] \) density where \( \Sigma = \begin{pmatrix} 1/2 \\ 1/2 \end{pmatrix} \). Let us initially consider the Type I error probabilities:

\[
P_{f_0} \left\{ \frac{L(f_2)}{L(f_1)} < 1; \frac{L(f_1)}{L(f_0)} > 1 \right\} = P_{f_0} \left\{ \frac{1}{n} \sum \frac{f_2(x_i)}{f_1(x_i)} < 0; \frac{1}{n} \sum \frac{f_1(x_i)}{f_0(x_i)} > 0 \right\}
\]

\[
= P_{f_0} \left\{ \frac{1}{n} \sum \frac{x_i}{2x_i - 2x_i} \leq -1; \frac{1}{n} \sum \frac{x_i}{2x_i - 2x_i} > 0 \right\}
\]

Let

\[
Y_1 = \frac{1}{2}x_2 - 2x_1 \\
Y_2 = \frac{1}{2}x_1 - x_2
\]

Then, under \( f_0 \)

\[
\begin{pmatrix} Y_1 \\ Y_2 \end{pmatrix} \sim N \left[ \begin{pmatrix} -3/2 \\ -1/2 \end{pmatrix}, \frac{1}{n} \begin{pmatrix} 7/4 & 7/8 \\ 7/8 & 7/2 \end{pmatrix} \right]
\]

In order to obtain random variables with mean zero the above probability may be rewritten:

\[
P_{f_0} \left\{ \frac{1}{n} \sum (Y_{1i} + \frac{3}{2}) \leq \frac{1}{2}; \frac{1}{n} \sum (Y_{2i} + \frac{1}{2}) \geq \frac{1}{2} \right\}
\]

In order to put the covariance matrix in the desired form we apply the
transformation

\[
\begin{pmatrix} z_1 \\ z_2 \end{pmatrix} = \begin{pmatrix} .664 & .470 \\ -.460 & -.326 \end{pmatrix} \begin{pmatrix} Y_1 + 3/2 \\ Y_2 + 1/2 \end{pmatrix}
\]

which yields

\[
\begin{pmatrix} z_1 \\ z_2 \end{pmatrix} \sim N \left( \begin{pmatrix} 0 \\ 1 \end{pmatrix}, \frac{1}{n} \right)
\]

or

\[
\sqrt{n} \begin{pmatrix} z_1 \\ z_2 \end{pmatrix} \sim N \left( \begin{pmatrix} 0 \\ 1 \end{pmatrix}, \frac{1}{1} \right)
\]

and the above probability is transformed into the probability of a wedge in the \((\sqrt{n} Z_1, \sqrt{n} Z_2)\) plane. The point in the wedge which is closest to the origin is \((.218 \sqrt{n}, -.311 \sqrt{n})\). Thus

\[
D = \sqrt{(.218 \sqrt{n})^2 + (-.311 \sqrt{n})^2}
\]

\[
= \sqrt{1.444n}
\]

and

\[
|D_1, D_2| = .067 \sqrt{n}
\]

so that, for large \(n\)

\[
P_f \left\{ \frac{L(f_2)}{L(f_1)} < 1; \frac{L(f_1)}{L(f_0)} > 1 \right\} \approx \beta_1(n) e^{-0.072n}
\]

Proceeding similarly we may approximate the other probabilities:

\[
P_f \left\{ \frac{L(f_2)}{L(f_1)} > 1; \frac{L(f_1)}{L(f_0)} > 1 \right\} \approx \beta_2(n) e^{-0.122n}
\]

\[
P_f \left\{ \frac{L(f_2)}{L(f_1)} > 1; \frac{L(f_1)}{L(f_0)} < 1 \right\} \approx \beta_3(n) e^{-0.163n}
\]
Using these we may write out the asymptotic expressions for \( R_1(\delta, j) \) and \( R_1(\delta_{NP, j+1, j}) \) for \( j = 0, 1, 2, \ldots \):

\[
R_1(\delta, 0) = \beta_3(n)e^{-1.63n} + \beta_4(n)e^{-2.19n} + \beta_5(n)e^{-0.72n} + \beta_2(n)e^{-1.22n}
\]

\[
R_1(\delta, 1) = \beta_5(n)e^{-1.59n} + \beta_6(n)e^{-3.47n} + \beta_7(n)e^{-0.72n} + \beta_2(n)e^{-1.22n}
\]

\[
R_1(\delta_{NP, 1, 0}) = \beta_7(n)e^{-0.76n} + \beta_4(n)e^{-2.19n} + \beta_8(n)e^{-2.23n} + \beta_1(n)e^{-0.72n}
\]

\[
R_1(\delta_{NP, 2, 1}) = \beta_5(n)e^{-1.59n} + \beta_9(n)e^{-1.67n} + \beta_2(n)e^{-1.22n} + \beta_{10}(n)e^{-1.63n}
\]

from which it's clear that, asymptotically

\[
R_1(\delta_{NP, 1, 0}) \geq R_1(\delta, 0) \geq R_1(\delta, 1) > R_1(\delta_{NP, 2, 1})
\]
This example is an illustration of a more general result:

**Theorem 3**: Suppose \( f_j(x_1, x_2) \) are bivariate normal distributions with distinct means \( \mu_{jj}, j = 0, 1, 2 \) and positive definite covariance matrix \( \Sigma \). Also suppose that

\[
(\mu_{00} - \mu_{11})(\Sigma^{-1}(\mu_{00} - \mu_{11})) 
\neq (\mu_{00} - \mu_{22})(\Sigma^{-1}(\mu_{00} - \mu_{22}))
\]  

where \( T \) is chosen so that \( T \Sigma T' = I \). Then

\[
\max_j R_j(\delta, j) \rightarrow \frac{\max_j R_j(\delta_{NP, j + 1}, j)}{R_j(\delta_{NP, j + 1}, 0)} \rightarrow 1 \quad \text{as} \quad n \rightarrow \infty .
\]

**Pf.**: The probabilities we must compute to evaluate \( R_1(\delta, j) \) and \( R_1(\delta_{NP, j + 1}, j) \) concern inequalities of the ratios \( \frac{L(f_1)}{L(f_2)} / \frac{L(f_1)}{L(f_0)} / \frac{L(f_2)}{L(f_0)} \). We shall denote the random variable \( X \) whose distribution is \( f \) by \( X | f \). We also define

\[
\begin{align*}
\mu_j &= T \mu_j, \quad j = 0, 1, 2 \\
Y &= TX
\end{align*}
\]

Then if \( X \) has distribution \( f_j(x) \)

\[
Y \sim N[\mu_j, I] \quad j = 0, 1, 2
\]

which we will denote by \( f_j(Y) \). Now

\[
\left. \frac{L(f_1)}{L(f_2)} \right| f_j = \exp\left[-\frac{1}{2} n \left( (x_i - \mu_{11})' \Sigma^{-1}(x_i - \mu_{11}) + \frac{1}{2} (x_i - \mu_{22})' \Sigma^{-1}(x_i - \mu_{22}) \right) \right| f_j
\]
Thus this likelihood ratio (and, by the same method, the other two) is equal to a likelihood ratio involving random variables whose bivariate normal distribution has covariance matrix $I$. This means that our problem has been reduced to the question of whether equation (18) holds when we are considering the densities $f_0', f_1'$ and $f_2'$. This has already been answered in the affirmative by Theorem 2 whenever $d(\mu_0', \mu_1') \neq d(\mu_0', \mu_2')$. Equation (17) is equivalent to this condition. q.e.d.

Our second example is a special case of Section D.1 of Chapter III in which $\delta$ is asymptotically optimal. Suppose $f_0(x), f_1(x)$ and $f_2(x)$ are three trinomial populations with probabilities $p_i$ for $f_0$, $q_i$ for $f_1$ and $r_i$ for $f_2$, $i = 1, 2, 3$. We have $\sum p_i = 3q_i = 3r_i = 1$ where e.g. $p_i$ is taken to be the probability of an event $g_i$ when $f_0(x)$ is the true distribution. If we consider $n$ trinomial trials and denote by $m_i$ the number of times $g_i$ occurred then $\sum m_i = n$. We may equivalently speak of the proportion of times the event $g_i$ occurred, $v_i = m_i/n$, $i = 1, 2, 3$. All points $v = (v_1, v_2, v_3)$ are members of the set

$$\Sigma = \{ v : v_i \geq 0, i = 1, 2, 3, \sum v_i = 1 \}$$

which is a simplex in $E_3$. One likelihood ratio in this case is:
with similar expressions being obtained for \( \frac{L(f_j)}{L(f_0)} \), \( j = 1,2 \). Thus the probabilities we are interested in computing are of the type: \[
P_f\{ \sum_{i} a_i v_i \geq 0; \sum_{i} b_i v_i \geq 0 \} \quad j = 0,1,2 \quad ,
\]
i.e., the same form as the probabilities which we approximated in Section D.1 of Chapter III.

Let us take a specific example with
\[
(p_1, p_2, p_3) = (2.2, 3.1, 5.0)
\]
\[
(q_1, q_2, q_3) = (6.0, 15.0, 25.0)
\]
\[
(r_1, r_2, r_3) = (5.0, 35.0, 15.0)
\]

Then if we calculate the magnitude of each term in \( R_1(\theta, j) \) and \( R_1(\theta_{NP, j+1}, j) \) \( j = 0,1 \) we obtain:

\[
R_1(\theta, 0) \sim 0.098 + 0.090 + 0.091 + 0.097
\]
\[
R_1(\theta, 1) \sim 0.087 + 0.092 + 0.091 + 0.097
\]
\[
R_1(\theta_{NP, 1}, 0) \sim 0.090 + 0.098 + 0.097 + 0.109
\]
\[
R_1(\theta_{NP, 2}, 1) \sim 0.087 + 0.103 + 0.091 + 0.095
\]
where the numbers on the right hand sides of the above equations represent the magnitudes of the bivariate probabilities composing this risk (see equations (8) and (9) in this section). Since the largest numbers denote the smallest magnitudes and since the two terms of magnitude .087 are identical

\[ R_1(\delta_{NP,1},0) \leq R_1(\delta,0) \leq R_1(\delta,1) \leq R_1(\delta_{NP,2},1) \sim .087 \]

or

\[ \lim_{n} \max_{j} R_1(\delta_{NP,j+1}, j) \leq \lim_{n} \max_{j} R_1(\delta_{NP,1}, j) \]

\[ = 1 \]

Our asymptotic considerations have reduced the small sample problem to a multiple decision problem, i.e., considering \( f_0, f_1 \) and \( f_2 \) but not mixtures of them. Considering which densities are used in computing the probabilities composing the risks we may rewrite \( R_1(\delta,j) \) as \( R_1(\delta,f_0^*, f_1^*) \). Then letting \( \Omega_\omega \) denote the set of 'distinct' densities contained in \( \mathcal{H}_A(\Omega_\omega = \{f_1, f_2\}) \) and letting \( \omega \) denote the set consisting of the single density which composes \( \mathcal{H}_0 \) we may write equation (3) in the form

\[ \max_{f_0 \in \omega} \max_{f_j \in \Omega_\omega} R_1(\delta, f_0^*, f_j) \]

\[ \lim_{n} \max_{f_0 \in \omega} \max_{f_j \in \Omega_\omega} R_1(\delta, f_0^*, f_j) = 1 \]

(19)

where \( \delta(f_0^*, f_j) \) is the Neyman-Pearson (or likelihood ratio) test of \( f_0 \) vs. \( f_j \). Now suppose that \( \omega \) contains \( I \) elements and \( \Omega_\omega \) contains \( J \) elements. We can repeat the procedure at the beginning of this section. First we define

\[ \tilde{\delta}(x_1, \ldots, x_n) \]
\[
\max_{f_i \in \Omega} \{ L(f_i) \} \\
\sim (x_1, \ldots, x_n) = 1 \iff \max_{g_i \in \Omega} \{ L(g_i) \} > 1
\]

Then we can expand \( R_1(\delta, g_i, f_j) \) and \( R_1(\delta(g_i, f_j), g_i, f_j) \) analogously to equations (8) and (9). There are now 2IJ terms in the expansions each of which can be written as the probability of each of \( I + J - 1 \) likelihood ratios being greater than 1; hence each is the probability that a certain \((I+J-1)\)-variate mean lies in the first orthant. Again any numerator (i.e., \( R_1(\delta, g_i, f_j) \) for any \( i \) and \( j \)) and any denominator (i.e., \( R_1(\delta(g_i, f_j), g_i, f_j) \) for any \( i \) and \( j \)) of the extension of equation (1) to this case have terms in common. The number of shared terms is 2IJ. We can use the multivariate generalizations of the bivariate methods of Chapter III to calculate these probabilities. If numerator and denominator have identical dominating terms then:

\[
\lim_{n \to \infty} \max_{g_i \in \Omega} \max_{f_j \in \Omega} R_1(\delta(g_i, f_j), g_i, f_j) = 1.
\]

B. Asymptotic Minimaxness in Exponential Families

Recalling the discussion in Section E of Chapter III with \( X \) convex and following Borovkov and Rogozin we define (for \( \Theta \in \Theta \)) \( A_{\Theta, L} \subset A \) by

\[
A_{\Theta, L} = \{ \Theta : I(\Theta, a) \leq L \}
\]

where \( I(\Theta, a) \) is given in Section E of Chapter III. For a set \( D \subset A \) we define:

\[
J_{\Theta, D} = \inf[ L : \mu(A_{\Theta, L} \cap D) > 0 ] > 0
\]
where $\mu$ is Lebesque measure.

Suppose (1) by convoluting $T$ enough times we obtain a random variable with bounded density (which is probably true) and (2) we have "$L^1$ order contact" between $D$ and the boundary of $A_{\theta, \tilde{D}}$ (in the sense of (2)).

Then, assuming the results of Borovkov and Rogozin,

$$P_{\tilde{T} \in \tilde{D} \cap \tilde{J}} = \tilde{\sigma}_0 n^{k/2 - 1 - \epsilon} \tilde{J}_{\theta, \tilde{D}}.$$  

(20)

Two simplifications of this formula are usually possible. The first replaces $\tilde{J}_{\theta, \tilde{D}}$ by

$$\tilde{J}'_{\theta, \tilde{D}} = \inf \{ L : A_{\theta, \tilde{L}} \cap \tilde{D} \neq \emptyset \}$$  

(21)

and the second replaces $\tilde{J}_{\theta, \tilde{D}}$ by

$$\tilde{J}''_{\theta, \tilde{D}} = \inf \{ l(\theta, \tilde{a}) \}$$  

(22)

for $\tilde{D} = \tilde{D} \cap \overline{CD}$ where $CD$ is the complement of $D$.

Note that, when $f(\theta)$ is in the closure of $CD$, version (22) of $J$ actually follows from (i.e., is equal to) version (21). This will be the case for the regions considered in connection with the likelihood ratio procedure below.

To show that versions (21) and (22) are equivalent we can show that

$$\inf_{\tilde{a} \in \tilde{D}} l(\theta, \tilde{a}) = \inf_{\tilde{a} \in \tilde{D}} l(\theta, \tilde{a}).$$  

This can be seen as follows: Suppose not. Then

$$\inf_{\tilde{a} \in \tilde{D}} l(\theta, \tilde{a}) < \inf_{\tilde{a} \in \tilde{D}} l(\theta, \tilde{a})$$

and = say $\inf_{\tilde{a} \in \tilde{D}} l(\theta, \tilde{a}) - \delta \equiv \tilde{I} - \delta$. By the continuity of $I(\theta, a)$ there is an $\tilde{a}^*$ in $D-\tilde{D}$ such that $l(\theta, a^*) = \tilde{I} - \frac{\delta}{2}$. On the other hand continuity also guarantees that there is an $a^*$ in the interior
of CD with \( l(\theta, a^*) = \gamma - \frac{A}{2} \), because \( l(\theta, f(\theta)) = 0 \) and

\[
\inf_{\theta \in \Theta} l(\theta, a) = \gamma. \tag{23}
\]

Thus \( A, \gamma, \frac{A}{2} \) contains \( a^* \) and \( \tilde{\gamma} \). Hence, since it is shown in (2) and (5) that the sets \( A, \gamma, \frac{A}{2} \) are convex, it follows that all points \( a \) on the line segment connecting \( a^* \) and \( \tilde{\gamma} \) have \( l(\theta, a) \geq \gamma - \frac{A}{2} \). But this contradicts equation (23).

We now proceed to indicate the asymptotic minimax nature of \( \tilde{\gamma} \) in the exponential case. We consider disjoint compact hypotheses (i.e., sets) \( H_0 \) and \( H_A \) in \( \Theta \) and critical regions \( A \) for \( \tilde{\gamma} \) in \( A \).

Our remarks, given in the form of three lemmas will pertain only to "smooth" \((H_0, H_A, A)\) in the sense that version \( J'' \) of equation (22) may be used in equation (20) for \( D = \Theta \) and all \( \theta \) in \( H_0 \) and in equation (20) for \( D = \Theta \) and all \( \theta \) in \( H_A \). We note that the test \( \tilde{\gamma} \) can be related to this situation by considering two elements \( \theta_0 \in H_0 \) and \( \theta_1 \in H_A \). Then

\[
l(\theta_0, a) \Rightarrow l(\theta_1, a) \Rightarrow a' \theta_0 - \psi(\theta_0) \Rightarrow a' \theta_1 - \psi(\theta_1)
\]

\[
\iff \frac{\partial}{\partial \theta}(x) \frac{f(\theta)}{f(\theta)} \overset{\sim}{=} 1.
\]

\( L_{\theta}(x) \) being proportional to the likelihood of \( \tilde{\gamma} \) when sampling from a density which is a member of the exponential family with parameter \( \theta_i \) \( i=0,1 \).

Thus the curve \( \tilde{\mathcal{R}} \) which we will refer to hence is the boundary of the critical region (and of the acceptance region) of the test \( \tilde{\gamma} \). Similarly the curve \( \mathcal{R} \) is the boundary of the critical region of the test \( \delta \).
Lemma 1: Consider the a-locus in A \( \sim \) = \( \{a: \min[\sim a' \theta - \psi(\theta)] = \min \sim a' \theta - \psi(\theta) \} \). Then \( \sim \) is also given by \( \sim = \{a: \min(\theta, \sim a) = \min(\theta, \sim a) \} \). {\( \sim a \in H_A \)}

\( \sim H_A \sim \)

Pf.: This follows because of the definition of \( l(\theta, \sim a) \), the compactness of \( H_0 \) and \( H_A \) and since for any two functions of \( \theta \), say \( f_1 \) and \( f_2 \)

\[
\min f_1(\theta) = \min f_2(\theta) \quad \min[f_1(\theta) + M] = \min[f_2(\theta) + M]
\]

where \( M \) is any constant. q.e.d.

Let \( \beta(a) = \min l(\theta, \sim a) = \min l(\theta, \sim a) \, \sim \in \sim \). Then

Lemma 2: Suppose that \( \inf \beta(a) \) is achieved at \( \sim \in \sim \). Let \( \sim \) be a smooth curve in A "separating" \( H_0 \) and \( H_A \). Then

\[
\inf l(\theta, \sim a) = \inf l(\theta, \sim a) = \beta(\sim a^*) \quad (24)
\]

Pf.: With regard to the equality in equation (24)

\[
\inf l(\theta, \sim a) = \inf \left[ \min l(\theta, \sim a) \right]
\]

\( \sim H_0 \cup H_A \)

\( \sim H_0 \cup H_A \)

\[
= \inf \left[ \min \{ \min l(\theta, \sim a), \, \min l(\theta, \sim a) \} \right]
\]

\( \sim H_0 \cup H_A \)

\( \sim H_0 \cup H_A \)

\[
= \inf \left[ \min \{ \beta(\theta), \, \beta(\theta) \} \right]
\]

\( \sim \)

by Lemma 1 and the definition of \( \beta(a) \). This quantity is the same as \( \inf \beta(a) \) which, by the statement of the Lemma, is \( \beta(a^*) \).

Define \( l_0(\theta) = \inf l(\theta, \sim a) \) with a similar definition of \( l_A(\sim) \). Also
define \( A_{\Omega_J} = \{ \mathbf{a}: l_0(\mathbf{a}) \leq J \} \) with \( A_{\Omega_A} \) similarly defined. Let us consider 
\( A_{\Omega_J} \beta(\mathbf{a}^*) \) and \( A_{\Omega_A} \beta(\mathbf{a}^*) \). By the definition of \( \mathbf{a}^* \) we can see that the only point (points) of intersection of \( A_{\Omega_J} \beta(\mathbf{a}^*) \) and \( A_{\Omega_A} \beta(\mathbf{a}^*) \) is (are) on the boundaries of the two sets, i.e., points \( \mathbf{a} \) such that \( l_0(\mathbf{a}) = l_{\Omega_A}(\mathbf{a}) = \beta(\mathbf{a}^*) \). If \( \mathcal{P} \) is any other curve "separating" \( \Omega_J \) and \( \Omega_A \) there are two possibilities: (1) \( \mathcal{P} \) runs through the interior of \( A_{\Omega_J} \beta(\mathbf{a}^*) \) or the interior of \( A_{\Omega_A} \beta(\mathbf{a}^*) \) or both. Let \( \mathbf{a} \) be any point in \( \mathcal{P} \) which is in the interior of 
\( A_{\Omega_J} \beta(\mathbf{a}^*) \) or the interior of \( A_{\Omega_A} \beta(\mathbf{a}^*) \). Since \( \mathbf{a} \) is a particular point of \( \mathcal{P} \)

\[
\inf_{\mathbf{a} \in \mathcal{P}} l(\mathbf{a}, \overline{a}) = \min \{ \inf_{\mathbf{a} \in \Omega_J} l_0(\mathbf{a}), \inf_{\mathbf{a} \in \Omega_A} l_{\Omega_A}(\mathbf{a}) \} = \min \{ l_0(\mathbf{a}), l_{\Omega_A}(\mathbf{a}) \}
\]

say. Thus \( \overline{a} \in A_{\Omega_J} \beta(\mathbf{a}^*) \). (If the min had been equal to \( l_{\Omega_A}(\mathbf{a}) \) then 
\( \overline{a} \in A_{\Omega_A} \beta(\mathbf{a}^*) \). We conclude that \( l_0(\overline{a}) \leq l_0(\mathbf{a}^*) = \beta(\mathbf{a}^*) \). (2) \( \mathcal{P} \) does not pass through the interior of \( A_{\Omega_J} \beta(\mathbf{a}^*) \) or of \( A_{\Omega_A} \beta(\mathbf{a}^*) \). From our remark about the intersection of \( A_{\Omega_J} \beta(\mathbf{a}^*) \) and \( A_{\Omega_A} \beta(\mathbf{a}^*) \) we see that \( \overline{a} \in \mathcal{P} \). There are three cases. First, if 

\[
\inf_{\mathbf{a} \in \mathcal{P}} l(\mathbf{a}, \overline{a})
\]

is achieved at \( \mathbf{a}^* \) then it is equal to 

\[
\inf_{\mathbf{a} \in \mathcal{P}} l(\mathbf{a}, \overline{a})
\]

Alternately if this inf is achieved at a point outside of both \( A_{\Omega_J} \beta(\mathbf{a}^*) \)
and $A_{H_A}^*\beta(\tilde{\omega})$ then it would be greater than $\beta(\tilde{\omega})$ which is a contradiction ($\tilde{\omega} \in \mathcal{A}$). Finally if the inf is achieved on the boundary of $A_{H_0}^*\beta(\tilde{\omega})$ or $A_{H_A}^*\beta(\tilde{\omega})$ then it is equal to $\beta(\tilde{\omega})$. q.e.d.

From Lemma 2 we immediately have that for any $\theta_0 \in H_0$, $\theta_1 \in H_A$

**Lemma 3:** \[
\inf_{\theta \in \mathcal{A}} \left[ \min \{ l(\theta_0, \theta), l(\theta_1, \theta) \} \right] \leq \inf_{\theta \in \mathcal{A}} \left[ \min \{ l(\theta_0, \theta), l(\theta_1, \theta) \} \right]
\]
for any "separating" curve $\mathcal{A}_0$.

If we have the finite $H_0$ and $H_A$ considered in Lemma 3 the probability of Type I error of a test $\delta$ of $H_0$ vs. $H_A$ may be written $\alpha_n(\delta, \theta_0)$ and the probability of Type II error may be written $\beta_n(\delta, \theta_1)$, $n$ denoting the sample size. Finally, returning to the $H_0$ and $H_A$ considered in Lemma 2 we can state the minimax property given there in terms of the risk $R_3$.

**Theorem 4** Consider any $\delta$ with smooth $\mathcal{A}$ in $A$. Given any pair $(\theta_0, \theta_1)$ $\theta_0^1 \in H_0$, $\theta_1^1 \in H_A$ there is a pair $(\theta_0^{11}, \theta_1^{11})$ $\theta_0^{11} \in H_0$, $\theta_1^{11} \in H_A$ such that:

\[
\lim_{n} \frac{\ln[\alpha_n(\delta, \theta_0^{11})] + \beta_n(\delta, \theta_1^{11})}{\ln[\alpha_n(\delta, \theta_0) + \beta_n(\delta, \theta_1)]} \geq 1.
\]

**Pf.:** From Lemma 3

\[
\inf_{\theta \in \mathcal{A}} \left[ \min \{ l(\theta_0, \theta), l(\theta_1, \theta) \} \right] \leq \inf_{\theta \in \mathcal{A}} \left[ \min \{ l(\theta_0, \theta), l(\theta_1, \theta) \} \right]
\]

Therefore given $(\theta_0^{11}, \theta_1^{11}) \exists (\theta_0^{1}, \theta_1^{1}) \exists$

\[
\inf_{\mathcal{A}} \min_{i} l(\theta_i, \theta) \leq \inf_{\mathcal{A}} \min_{i} l(\theta_i, \theta)
\]

Thus from equation (20) we have the desired result. q.e.d.
V. BIBLIOGRAPHY


VI. ACKNOWLEDGEMENTS

I wish to express my gratitude to Professor H. T. David whose advice and guidance during the preparation of this thesis were invaluable.