ASYMPTOTICALLY OPTIMAL TESTS FOR MULTIVARIATE NORMAL DISTRIBUTIONS

By David G. Herr

University of North Carolina, Charlotte

1. Introduction. Early work on probabilities of large deviations had to do with probabilities of deviations of the sample mean as the sample size increased. The papers of Chernoff [3] and Bahadur and Rao [1] are notable in this area. More general sets were considered by Borokov and Rogozin [2], Hoeffding [4], and Sanov [7]. Sanov [7] proved a result concerning probabilities of sets contained in the space of maximum likelihood estimates of the parameters of a multinomial distribution as the sample size increased. His result was of consequence only when the true parameter point was sufficiently "far" from the set in question. Hoeffding [5] sharpened this result to the following. If Ω is defined by

$$\Omega = \{(x_1 \cdots, x_k) : x_1 \geqq 0, \cdots, x_k \geqq 0; x_1 + \cdots + x_k = 1\}$$

and $z^{(N)}$ as the k-vector of maximum likelihood estimates of the k-vector $\mathbf{p}$ of parameters for a k dimensional multinomial distribution based on a sample size N; then the probability $P_N(A \mid \mathbf{p})$ that $z^{(N)} \varepsilon A \subset \Omega$ is given by

$$P_N(A \mid \mathbf{p}) = \exp[-NI(A^{(N)}, \mathbf{p}) + O(\log N)]$$

where

$$I(A^{(N)}, \mathbf{p}) = \inf\{I(\mathbf{x}, \mathbf{p}) : \mathbf{x} \varepsilon A^{(N)}\}$$

for

$$I(\mathbf{x}, \mathbf{p}) = \textstyle\sum_{i=1}^{k} x_i \log(x_i/p_i),$$

if $\mathbf{p} = (p_1, \cdots, p_k)$, $\mathbf{x} = (x_1, \cdots, x_k)$; uniformly in A and $\mathbf{p}$. Here $A^{(N)}$ is the intersection of the space of maximum likelihood estimates, $z^{(N)}$, with A.

Applying this estimate to the error probabilities involved in testing simple and composite hypotheses, Hoeffding was able to substantiate the following proposition: "If a given test of size α_N is 'sufficiently different' from a likelihood ratio test, then there is a likelihood ratio test of size $\leqq \alpha_N$ which is considerably more powerful than the given test at 'most' points $\mathbf{p}$ in the set of alternatives when N is large enough, provided that $\alpha_N \to 0$ at a suitable rate."

This result depends almost entirely on the probability estimate. It is conjectured by Dr. Hoeffding that such an estimate holds for a wide class of distributions of exponential type. The present work is an attempt to extend the probability result to the class of non-singular multivariate normal distributions and to apply this result in a way analogous to that of Hoeffding [5] to the problem of comparing tests of simple and composite hypotheses with appropriate likeli-

Received 2 March 1967; revised 23 June 1967.

hood ratio tests in the multivariate normal case. As Hoeffding points out in his paper Wald [8] has shown that the likelihood ratio test (LRT) has certain asymptotically optimal properties when the error probabilities are bounded away from zero, but the present results are of a different nature.

The principal barrier to an easy extension is the key probability estimate. In the multinomial case the probability of any non-empty set in the space of maximum likelihood estimates for the parameters is bounded below by the probability of any singleton subset, which is positive since the multinomial distribution is discrete. This fact leads to a uniform lower bound for the probability of any such non-empty set and from this the desired probability estimate is easily obtained. In the multivariate normal case however the distribution of the maximum likelihood estimate (MLH) of p is continuous so that non-empty sets may have probability as close to zero as desired and no positive uniform lower bound exists. The problem then is to restrict attention to a class of sets which is relatively large and for which a desirable lower bound exists. We have chosen to consider all measurable sets which contain a "sphere" with radius depending on N and satisfying some additional restrictions. This choice is somewhat arbitrary and was dictated by the simplification which resulted.

It is shown that for such a class of sets, $B_0(p)$, in the space of maximum likelihood estimates of the parameter $p = (u, A)$, u the mean vector and A the variance-covariance matrix, of a multivariate normal distribution; the probability $P_N(B \mid p)$ of a set B in the class $B_0(p)$ satisfies

$$P_N(B \mid p) = \exp\left[-N(1 + O(N^{-1}))J(B, p) + O(\log N)\right]$$

uniformly in B and p. Here $J(B, p) = \inf \{J(\dot{p}, p) : \dot{p} \varepsilon B\}$ for

$$J(\dot{p}, p) = \tfrac{1}{2}[\operatorname{tr} \dot{A}A^{-1} - k - \log |\dot{A}A^{-1}| + (\dot{u} - u)'A^{-1}(\dot{u} - u)]$$

if $\dot{p} = (\dot{u}, \dot{A})$ is the maximum likelihood estimate of p. Note that the result only has significance if $J(B, p) > 0$ which implies that p is "far" from B. In this sense we are dealing with probabilities of large deviations.

This probability estimate is then used to find sufficient conditions for substantiating Hoeffding's proposition in the case of tests concerning hypotheses about the parameter $p = (u, A)$ of a multivariate normal distribution. The set of alternatives to the hypothesis $p \varepsilon H_0$ for which a likelihood ratio test is not "sufficiently different" from a given test is considered for a wide class of sets H_0.

An example is considered in which it is shown that there is a likelihood ratio test which is asymptotically better at "most" alternatives than a test of S. N. Roy's [6] for the composite hypothesis $A = I$, the k by k identity matrix, versus $A \neq I$ in the case $k = 2$, in the sense that the likelihood ratio test has smaller size and the ratio of the type II error probability of the likelihood ratio test to that of Roy's test tends to zero faster than any power of N.

It is of interest to note that these results supply partial answers to questions concerning the extension of Hoeffding's results to distributions other than the multinomial raised by Chernoff in his discussion of Hoeffding's paper [5] and further reinforce the ideas expressed by Chernoff concerning the "distance" interpretation of the function J. As is glaringly brought out in the example and elsewhere, the class $B_0(p)$ is too complex and limited for these results to be an entirely satisfactory extension of Hoeffding's work. However, the work here does shed some light on the problems encountered in such an extension.

Before continuing a comment on the notation used will be helpful. In the k-variate normal case the MLH estimate of the mean vector and of the variance-covariance matrix are respectively a k dimensional vector and a real, symmetric, almost everywhere positive definite, k by k matrix [8]. The notation used is respectively $\dot{u}$ and $\dot{A}$. However, when considering the pair $(\dot{u}, \dot{A})$ as variables or points in a space, we are actually thinking of the $n = (k(k+1)/2) + k$ dimensional vector obtained from $(\dot{u}, \dot{A})$. If $\dot{u} = (t^1, \cdots, t^k)$ and $\dot{A} = (t_{ij})$, $1 \leqq k$; this vector is

$$(1.1) \qquad (t^1, \cdots, t^k, t_{11}, t_{12}, \cdots, t_{1k}, t_{22}, t_{23}, \cdots, t_{2k}, t_{33}, \cdots, t_{kk}).$$

In the analysis of the problem at hand the notation $(\dot{u}, \dot{A})$ is by far the most convenient since the function J depends on the vector in (1.1) only through $\dot{u}$ and $\dot{A}$. For this reason we adopt the convention that whenever possible we shall use the notation $\dot{p} = (\dot{u}, \dot{A})$.

It is hoped that these few comments will make it possible to read the following without the need of two parallel and necessarily cumbersome notations.

2. Probabilities of large deviations in multivariate normal distributions. Let $x_1, x_2, \cdots, x_N$ be independent, identically distributed k by 1 real vectors from a k dimensional multivariate normal distribution with mean vector u and positive definite (pd) variance-covariance matrix A. Define the space Q_0 by

$$(2.1) \quad Q_0 = \{(u, A): u \text{ is a } k \text{ by } 1 \text{ real vector, } A \text{ is a } k \text{ by } k \text{ real, symmetric, pd matrix}\}.$$

Define a function J from $Q_0 \times Q_0$ into the real line by

$$(2.2) \qquad J(p_1, p_2) = \int_{E_k} \log [f_{p_1}(x)/f_{p_2}(x)] f_{p_1}(x)\, dx$$

where $p_1, p_2 \,\varepsilon\, Q_0$, E_k is Euclidean k-space, and f_{p_i} is the density with respect to Lebesgue measure of the k dimensional multivariate normal probability measure with parameter p_i, $i = 1, 2$. It follows that

$$(2.3) \quad J(p_1, p_2) = \tfrac{1}{2}[\operatorname{tr} A_1A_2^{-1} - k - \log |A_1A_2^{-1}| + (u_1 - u_2)'A_2^{-1}(u_1 - u_2)]$$

where for any k by k matrix A, $|A|$ = determinant of A, and $\operatorname{tr} A$ = trace of A. Also $p_i = (u_i, A_i)$, $i = 1, 2$.

If $\dot{u}$ and $\dot{A}$ are the maximum likelihood estimates of the mean vector u and the variance-covariance matrix A respectively of a k-variate normal distribution, then

$$\dot{u} = N^{-1} \textstyle\sum_{i=1}^N x_i, \qquad \dot{A} = N^{-1} \textstyle\sum_{i=1}^N (x_i - \dot{u})(x_i - \dot{u})'$$

and

(2.4) $J(\dot{p}, p) = \frac{1}{2}(\operatorname{tr} \dot{A}A^{-1} - k - \log|\dot{A}A^{-1}| + (\dot{u} - u)'A^{-1}(\dot{u} - u)]$

for $\dot{p} = (\dot{u}, \dot{A})$, $p = (u, A)$. For any measurable set $B \subset Q_0$ define $J(B, p)$ by

(2.5) $$J(B, p) = \inf\{J(\dot{p}, p) : \dot{p} \,\varepsilon\, B\}.$$

We are now able to state the following theorem.

THEOREM 2.1. *For any Lebesgue measurable set $B \subset Q_0$ and any $p \,\varepsilon\, Q_0$*

(2.6) $$P_N(B \mid p) \leqq \exp[-NJ(B, p) + (2k + 3)J(B, p) + O(\log N)]$$

uniformly in B, p as N tends to infinity. In fact for N sufficiently large

(2.7) $$P_N(B \mid p) \leqq C_1 N^{k(k+1)/2} \exp[-NJ(B, p) + (2k + 3)J(B, p)]$$

uniformly in B, p where $C_1 > 0$ is a constant.

The proof of this theorem follows from the relation

(2.8) $$P_N(B \mid p) \leqq P_N(\{\dot{p} : J(\dot{p}, p) \geqq J(B, p)\} \mid p) \leqq E_p[\exp(t(J(\dot{p}, p) - J(B, p)))],$$

for $t \geqq 0$ by a series of relatively uninteresting calculations.

In order to obtain a corresponding lower bound for $P_N(B \mid p)$ we shall need the density of $\dot{p}$ which is given by

(2.9) $$C_N |\dot{A}|^{-(k+2)/2} \exp[-NJ(\dot{p}, p)], \quad C_N = \exp[O(\log N)].$$

Thus

(2.10) $$P_N(B \mid p) = \int_B |\dot{A}|^{-(k+2)/2} C_N \exp[-NJ(\dot{p}, p)]\, d\dot{p},$$

where

$$C_N = \pi^{-k(k+1)/4}[\textstyle\prod_{i=1}^{k} \Gamma((N - i)/2)]^{-1} \exp[(k/2)N(\log(N/2) - 1)]$$

and $C_N = \exp[O(\log N)]$ uniformly in p as N tends to infinity.

We shall need the following definition.

DEFINITION 2.1. The essential infimum of $J(\dot{p}, p)$ over a set $B \subset Q_0$ will be defined by

$$\operatorname{ess}_B \inf J(\dot{p}, p) = \sup\{c : m(\{\dot{p} : J(\dot{p}, p) < c\} \cap B) = 0\}$$

where m is $(k + k(k + 1)/2) = n$ dimensional Lebesgue measure.

We shall denote $\operatorname{ess}_B \inf J(\dot{p}, p)$ by $c_B(p)$. Since

$$m(\{\dot{p} : J(\dot{p}, p) < J(B, p)\} \cap B) = 0,$$

it follows that $c_B(p) \geqq J(B, p)$. We may, by excluding a set of measure zero from B, insure that $c_B(p) = J(B, p)$ and from now on we shall assume this has been done and that $c_B(p) = J(B, p)$. Such an adjustment will be termed adjusting for the ess inf. The proof of these remarks will be given in Section 4.

Let us attempt to find a lower bound for $P_N(B \mid p)$ of the same form as that of the upper bound in (2.7). Toward this end define the set $D_\delta(p)$ by

$$D_\delta(p) = \{\dot{p}: c_B(p) + \delta < J(\dot{p}, p) < c_B(p) + 2\delta\} \tag{2.11}$$

for $\delta \geqq 0$. Let $\{\delta_N\}$ be a sequence of positive terms with limit zero. Suppose for each N there exists a set $B_{\delta_N}(p)$ such that

$$P_N(B_{\delta_N}(p) | p) \geqq C_0 N^r \exp[-(N + s)c_B(p)] \tag{2.12}$$

for (C_0, r, s) independent of (N, p) and $C_0 > 0$. Further suppose for each N

$$B_{\delta_N}(p) \subset D_{\delta_N}(p) \cap B.$$

Then $P_N(B \mid p) \geqq P_N(B_{\delta_N}(p) \mid p)$ and we would have a lower bound for $P_B(B \mid p)$ of the desired form.

In order to investigate the hypothesized sets $B_{\delta_N}(p)$ further we shall attempt to find a lower bound for the diameter of the set $D_\delta(p)$, $\delta > 0$. For $\dot{p} = (\dot{u}, \dot{A})$ define

$$\|\dot{p}\| = [\textstyle\sum_{i=1}^k (t^i)^2 + \sum_{1 \leqq i,j \leqq k} (t_{ij})^2]^{\frac{1}{2}} \tag{2.13}$$

where $\dot{u}' = (t^1, \cdots, t^k)$ and $\dot{A} = (t_{ij})$, $i, j = 1, \cdots, k$. Thus we wish to find a lower bound for $\|\dot{p}_1 - \dot{p}_2\|$ if $J(\dot{p}_1, p) = c_1$, $J(\dot{p}_2, p) = c_2$, and $c_2 > c_1 \geqq 0$.

Let $p = (u, A)$ and S be the lower triangular matrix with positive diagonal terms such that $SS' = A$. Then to facilitate matters we make the transformation τ_p from Q_0 into itself defined by

$$\tau_p(\dot{p}) = (S^{-1}(\dot{u} - u), S^{-1}\dot{A}S'^{-1}) = (v, V) = q, \tag{2.14}$$

say, and note

$$J(\dot{p}, p) = \tfrac{1}{2}[\operatorname{tr} V - k - \log |V| + v'v] = J(q, (0, I)). \tag{2.15}$$

Define J^0 on Q_0 by

$$J^0(q) = J(q, (0, I)). \tag{2.16}$$

Now if $\lambda_1(A) \leqq \cdots \leqq \lambda_k(A)$ are the ordered eigenvalues of A, if

$$b_1^2 = [\textstyle\sum_{i=1}^k (\lambda_i(A))^{-1}]^2 + \sum_{i=1}^k (\lambda_i(A))^{-1},$$

$$b_2^2 = [\textstyle\sum_{i=1}^k \lambda_i(A)]^2 + \sum_{i=1}^k \lambda_i(A),$$

and if $q_i = \tau_p(\dot{p}_i)$, $i = 1, 2$; then

$$b_1^2 \|\dot{p}_1 - \dot{p}_2\|^2 \geqq \|q_1 - q_2\|^2, \quad \|\dot{p}_1 - \dot{p}_2\|^2 \leqq b_2^2 \|q_1 - q_2\|^2. \tag{2.17}$$

Since $J(\dot{p}, p) = J^0(\tau_p(\dot{p})) = J^0(q)$ we thus seek a lower bound for $\|q_1 - q_2\|$ subject to $J^0(q_1) = c_1$, $J^0(q_2) = c_2$, $c_2 > c_1 \geqq 0$. If $q = (v, V)$ and $v' = (v^1, \cdots, v^k)$, $V = (v_{ij})$, $i, j = 1, \cdots, k$; then let $\nabla J^0(q)$ denote the gradient of J^0 with respect to the n dimensional vector

$$(v^1, \cdots, v^k, v_{11}, v_{12}, \cdots, v_{1k}, v_{22}, v_{23}, \cdots, v_{2k}, \cdots, v_{kk}). \tag{2.18}$$

Since J^0 is convex (see Section 4),

$$0 > c_1 - c_2 = J^0(q_1) - J^0(q_2) \geqq (q_1 - q_2)'\nabla J^0(q_2) \geqq -\|q_1 - q_2\| \, \|\nabla J^0(q_2)\|$$

and thus

$$c_2 - c_1 \leqq \|q_1 - q_2\| \, \|\nabla J^0(q_2)\|. \tag{2.19}$$

If we let $b = \sup \{\|\nabla J^0(q)\| : J^0(q) = c_2\}$, we have, assuming $b > 0$,

$$\|q_1 - q_2\| \geqq (c_2 - c_1)/b. \tag{2.20}$$

If a^{ij} is the cofactor of v_{ij}, then the components of $\nabla J^0(q)$ are given by

$$\partial J^0(q)/\partial v^i = v^i, \qquad \partial J^0(q)/\partial v_{ij} = -a^{ij}/|V|, \qquad i \neq j, \tag{2.21}$$
$$= \tfrac{1}{2}(1 - a^{ii}/|V|), \qquad i = j.$$

It follows that

$$\|\nabla J^0(q)\|^2 = \|v\|^2 + \tfrac{1}{4}(1 - a^{11}/|V|)^2 + (a^{21}/|V|)^2 + \cdots + \tfrac{1}{4}(1 - a^{kk}/|V|)^2$$

where $\|v\|^2 = \sum_{i=1}^k (v^i)^2$. Now to find an upper bound for b note that

$$\begin{aligned} \|\nabla J^0(q)\|^2 &\leqq (1 - a^{11}/|V|)^2 + 2(a^{21}/|V|)^2 \\ &\quad + \cdots + (1 - a^{kk}/|V|)^2 + \|v\|^2 \\ &= \|I - V^{-1}\|^2 + \|v\|^2 \\ &= \textstyle\sum_{i=1}^k (\lambda_i(V)^{-1} - 1)^2 + \|v\|^2 \end{aligned} \tag{2.22}$$

where for any real k by k matrix $D = (d_{ij})$, $\|D\|^2 = \sum_{i=1}^k \sum_{j=1}^k (d_{ij})^2$. We know from (2.16) that if $J^0(q) = c_2$ then q satisfies

$$\textstyle\sum_{i=1}^k (\lambda_i(V) - 1 - \log \lambda_i(V)) + \|v\|^2 = 2c_2 . \tag{2.23}$$

Now if λ is any real number in the interval $(0, 1)$ and $c' > 0$, then using the fact that $\log(1 + \lambda) < \lambda$ it is readily seen that $\lambda - 1 - \log \lambda \leqq c'$ implies that $\exp(c' + 1) - 1)^{-1} < \lambda$. Likewise if λ is greater than 1 and $c' > 0$, then it follows from $\log \lambda = 2 \log \lambda^{\frac{1}{2}} < 2(\lambda^{\frac{1}{2}} - 1)$ that $\lambda - 1 - \log \lambda \leqq c'$ implies $\lambda < ((c')^{\frac{1}{2}} + 1)^2$. So that (2.23) implies that

$$(\exp(c' + 1) - 1)^{-1} < \lambda_i(V) < ((c')^{\frac{1}{2}} + 1)^2, c' = 2c_2 - \|v\|^2. \tag{2.24}$$

Notice that we have made use of the fact that $\lambda - 1 - \log \lambda \geqq 0$ for all $\lambda > 0$, and that V is positive definite. By using (2.22) and by performing some elementary calculations, we are able to establish that

$$\|\nabla J^0(q)\|^2 < 9k \exp(4c_2 + 2). \tag{2.25}$$

Thus the right hand side of (2.25) is an upper bound for b^2 and when substituted in (2.20) yields

$$\|q_1 - q_2\| > (c_2 - c_1)/(3k^{\frac{1}{2}} \exp(2c_2 + 1)). \tag{2.26}$$

If we restrict our attention to those p's in Q_ϵ which is defined by

(2.27) $$Q_\epsilon = \{p: \lambda_1(A) \geqq \epsilon\} \quad \text{where} \quad \epsilon > 0;$$

then (2.17) together with (2.26) implies that

(2.28) $$\|\dot{p}_1 - \dot{p}_2\| \geqq [\epsilon'/3k^{\frac{3}{2}}](c_2 - c_1) \exp(-2c_2 - 1), \epsilon' = \epsilon/(1 + \epsilon)^{\frac{1}{2}},$$

for $J(\dot{p}_1, p) = c_1$, $J(\dot{p}_2, p) = c_2$, $p \in Q_\epsilon$. Moreover if $c_1 = c_B(p) + \delta$, $c_2 = c_B(p) + 2\delta$, then (2.28) implies

(2.29) $$\|\dot{p}_1 - \dot{p}_2\| \geqq \delta(\epsilon'/3k^{\frac{3}{2}}) \exp(-2c_B(p) - 4\delta - 1).$$

For $0 < c_1 < c_2$ we have the following lemma.

LEMMA 2.1. *If $\dot{p}_3$ satisfies $J(\dot{p}_3, p) = (c_1 + c_2)/2$ and $p \in Q_\epsilon$, then $S_2(\dot{p}_3)$ defined by*

$$S_2(\dot{p}_3) = \{\dot{p}: \|\dot{p} - \dot{p}_3\| < ((c_2 - c_1)/6k^{\frac{3}{2}})\epsilon' \exp(-2c_2 - 1)\}$$

satisfies $S_2(\dot{p}_3) \subset \{\dot{p}: c_1 < J(\dot{p}, p) < c_2\}$.

For the case $c_1 = c_B(p) + \delta$, $c_2 = c_B(p) + 2\delta$ we have then that every point in $\{\dot{p}: J(\dot{p}, p) = c_B(p) + (\frac{3}{2})\delta\}$ is the center of a hypersphere of radius δ' where

(2.30) $$\delta' = \delta(\epsilon'/6k^{\frac{3}{2}}) \exp(-2c_B(p) - 4\delta - 1), \epsilon' = \epsilon/(1 + \epsilon)^{\frac{1}{2}},$$

and each such hypersphere is contained in $D_\delta(p)$ for $p \in Q_\epsilon$. Suppose $D_\delta(p) \cap B$ contains one such sphere, say S_0, then

$$P_N(B \mid p) \geqq P_N(S_0 \mid p) = \int_{S_0} |\dot{A}|^{-(k+2)/2} C_N \exp[-NJ(\dot{p}, p)]\, d\dot{p}.$$

To conveniently handle this integral we could make the transformation τ_p, but a simpler approach will be to write $|\dot{A}| = |V|\,|A|$ and then

$$\int_{S_0} |\dot{A}|^{-(k+2)/2} C_N \exp[-NJ(\dot{p}, p)]\, d\dot{p}$$
$$\geqq (\max_{S_0} |V|)^{-(k+2)/2} |A|^{-(k+2)/2} C_N \exp[-Nc_B(p) - N2\delta] m(S_0).$$

Now

$$\max_{S_0} |V| \leqq \max_{D_\delta(p)} |V| = \max_{D_\delta(p)} \prod_{i=1}^{k} \lambda_i(V) < [(c_B(p) + 2\delta)^{\frac{1}{2}} + 1/2^{\frac{1}{2}}]^{2k} 2^k$$

since $J^0(q) = J(\dot{p}, p)$ and in view of the inequality (2.24). Thus

(2.31) $$P_N(B \mid p) \geqq [(c_B(p) + 2\delta)^{\frac{1}{2}} + 1/2^{\frac{1}{2}}]^{-k(k+2)} |A|^{-(k+2)/2} C_N \cdot 2^{-k(k+2)/2} \exp[-N(c_B(p) + 2\delta)] m(S_0).$$

Now $m(S_0)$ is given by

(2.32) $$m(S_0) = c_0(\delta')^n, \ n = k + k(k + 1)/2, \ c_0 \text{ a constant.}$$

Suppose $p \in Q^M$ where

(2.33) $$Q^M = \{p: \lambda_k(A) \leqq M\}, \ M \text{ a real positive constant.}$$

Then $|A| \leqq M^k$ and $P_N(B \mid p)$ is bounded below by

(2.34) $$C_N(\epsilon, M, k)\delta^n \exp[-(N + k(k + 2) + 2n)(c_B(p) + 2\delta)]$$

where

$$C_N(\epsilon, M, k) = C_N c_0 (4M)^{-k(k+2)/2} (\epsilon'/e6k^{\frac{3}{2}})^n. \tag{2.35}$$

If $\delta = (M_1/N) \log N$, then the lower bound for $P_N(B \mid p)$ in (2.34) becomes

$$\exp[-Nc_B(p) - (k(k+2) + 2n)c_B(p) + O(\log N)] \tag{2.36}$$

In order to summarize the preceding discussion let us define $B_0(p)$ as the class of sets B in E_n such that

(i) B is Lebesgue measurable,

(ii) there is a $\Delta_0 > 0$ such that for $0 < \delta < \Delta_0$, $D_\delta(p) \cap B$ contains a hypersphere of radius δ' as given in (2.30).

Then we have proved the following lemma.

LEMMA 2.2. *If $B \in B_0(p)$ and if $p \in Q_\epsilon \cap Q^M$, then*

$$P_N(B \mid p) \geqq \exp[-Nc_B(p) - (k(k+2) + 2n)c_B(p) + O(\log N)] \tag{2.37}$$

and the $O(\log N)$ term is the same for all $p \in Q_\epsilon \cap Q^M = \{p: \epsilon \leqq \lambda_1(A), \lambda_k(A) \leqq M\}$.

It follows that for N sufficiently large there exist constants $C_0 > 0$, r, s, such that if $p \in Q_\epsilon \cap Q^M$ and $B \in B_0(p)$, then

$$P_N(B \mid p) \geqq C_0 N^r \exp[-(N + s)c_B(p)] \tag{2.37}$$

uniformly in B and p. Thus we have the following theorem.

THEOREM 2.2. *If $p \in Q_\epsilon \cap Q^M$ and $B \in B_0(p)$, then for sufficiently large N*

$$C_0 N^r \exp[-(N + s)c_B(p)] \leqq P_N(B \mid p) \leqq C_1 N^{k(k+1)} \exp[-(N - s)c_B(p)] \tag{2.38}$$

for some constants $C_0 > 0$, $C_1 > 0$, r, and s uniformly in B and p.

3. The role of the likelihood ratio test. We would now like to apply the results of Theorem 2.2 to the problem of comparing tests of the hypothesis $p \in H_0 \subset Q_0$. We wish to compare a sequence of arbitrary, non-randomized tests whose acceptance and rejection regions are in $B_0(p)$ for the pertinent p, with a certain sequence of likelihood ratio tests (LRT).

Using the density of $\dot{p}$, it is readily verified that in the multivariate normal case the LRT for testing $p \in H_0$ versus $p \notin H_0$ based on observing the maximum likelihood estimate $\dot{p}$ of p with sample size N rejects the hypothesis when $J(\dot{p}, H_0)$ exceeds a constant. Here $J(\dot{p}, H_0)$ is given by

$$J(\dot{p}, H_0) = \inf\{J(\dot{p}, p): p \in H_0\}. \tag{3.1}$$

Let

$$B_N = \{\dot{p}: J(\dot{p}, H_0) \geqq c_N\} \tag{3.2}$$

for $c_N > 0$ and suppose B_N and B_N', the rejection and acceptance regions of a LRT respectively, are in $B_0(p)$. Then for $p \in Q_\epsilon \cap Q^M$

$$P_N(B_N \mid p) = \exp[-N(1 + O(N^{-1}))c_{B_N}(p) + O(\log N)] \tag{3.3}$$

and if the size of the test defined by B_N is $\alpha(B_N)$, then

$$(3.4)\quad \alpha(B_N) = \sup \{P_N(B_N \mid p) : p \varepsilon H_0\} = \exp[-N(1 + O(N^{-1}))c_N + O(\log N))]$$

where $\inf_{H_0} \inf_{B_N} J(\dot{p}, p) = \inf_{B_N} \inf_{H_0} J(\dot{p}, p) = c_N$.

Let a sequence of arbitrary tests of the hypothesis $p \varepsilon H_0$ be represented by the respective rejection regions which we shall call A_N. Suppose A_N and A_N' are in $B_0(p)$ for N sufficiently large and for pertinent p. Then if the size of the test A_N is given by $\alpha(A_N)$,

$$(3.5)\quad \alpha(A_N) = \sup \{P_N(A_N \mid p) : p \varepsilon H_0\} = \exp[-N(1 + O(N^{-1}))J(A_N, H_0) + O(\log N)]$$

where

$$(3.6)\quad J(A_N, H_0) = \inf_{A_N} \inf_{H_0} J(\dot{p}, p) = \inf_{H_0} \inf_{A_N} J(\dot{p}, p).$$

If we choose $c_N = J(A_N, H_0)$, then $B_N \supset A_N$ and the LRT B_N is at least as powerful as A_N. Since $B_N \supset A_N$, $\alpha(B_N) \geqq \alpha(A_N)$. However, if we assume

$$(3.7)\quad Nc_N / \log N \to \infty \quad \text{as} \quad N \to \infty,$$

then $\alpha(A_N)$ will tend to zero faster than any power of N and $(\log N)/\log \alpha(A_N)$ will tend to zero as $N \to \infty$. Thus if $c_N = O(\log N)$,

$$\log \alpha(B_N) = \log \alpha(A_N) + O(\log N)$$

which implies that

$$(3.8)\quad \log \alpha(B_N)/\log \alpha(A_N) = 1 + o(1)$$

so that in this sense the size of B_N is approximately that of A_N.

It is of interest to note that B_N not only contains A_N, but is actually equal to the union of all critical regions A_N^* such that $J(A_N^*, H_0) \geqq c_N$. Furthermore if we consider $H_0 = \{p_0\}$ then B_N is the union of critical regions of most powerful tests of the simple hypothesis $p = p_0$ against simple alternatives. Similarly for general H_0, B_N is the intersection over H_0 of the union of critical regions of most powerful tests of the simple hypothesis $p = p_0$ against simple alternatives. These facts may be used to explain the asymptotically optimal character of LRT.

As we have mentioned $\alpha(A_N) \leqq \alpha(B_N)$. However, we would like to find a LRT of $p \varepsilon H_0$ whose size is less than or equal to that of A_N and whose power is still better in some sense. Thus consider

$$(3.9)\quad B_N^+ = \{\dot{p} : J(\dot{p}, H_0) \geqq c_N + \Delta_N\}$$

and require $\alpha(B_N^+) \leqq \alpha(A_N)$. If we use the result of Theorem 2.2, this requirement becomes

$$(3.10)\quad \begin{aligned} \log \alpha(B_N^+) &\leqq -N(c_N + \Delta_N) + s(c_N + \Delta_N) + K_2 \log N \\ &\leqq -Nc_N - sc_N + K_1 \log N \leqq \log \alpha(A_N) \end{aligned}$$

where

$$K_1 = r + (\log C_0)/\log N, \qquad K_2 = k(k+1)/2 + (\log C_1)/\log N.$$

We see at once that (3.10) is implied by

$$(3.11) \qquad \Delta_N = (K_2 - K_1)(\log N)/(N - s) + 2sc_N/(N - s);$$

for $c_N = O(\log N)$, $N\Delta_N = O(\log N)$ as $N \to \infty$. Also from Theorem 2.2 it follows that $\Delta_N \geqq 0$ and (3.11) implies $\Delta_N \to 0$ as $N \to \infty$ for $c_N = O(\log N)$. Thus for Δ_N as defined in (3.11) we have

$$(3.12) \quad \exp[-N(1 + O(N^{-1}))c_N - O(\log N)] = \alpha(B_N^+) \leqq \alpha(A_N) = \exp[-N(1 + O(N^{-1}))c_N + O(\log N)].$$

Now if we apply the result of Theorem 2.2 to $B_N^{+\prime}$ and A_N', we have for the probabilities of falsely accepting the hypothesis that $p \varepsilon H_0$, i.e., $p \not\varepsilon H_0$,

$$(3.13) \quad P_N(B_N^{+\prime} \mid p) = \exp[-N(1 + O(N^{-1}))c_{B_N^{+\prime}}(p) + O(\log N)],$$

$$(3.14) \quad P_N(A_N' \mid p) = \exp[-N(1 + O(N^{-1}))c_{A_N'}(p) + O(\log N)],$$

for the tests B_N^+ and A_N respectively. If both of these probabilities are positive we have

$$(3.15) \quad P_N(B_N^{+\prime} \mid p)/P_N(A_N' \mid p) \leqq \exp[-N(d_N(p) - e_N(p) - sf_N(p)/N) + O(\log N)]$$

where

$$(3.16) \quad \begin{aligned} d_N(p) &= J(B_N', p) - J(A_N', p) \geqq 0, \\ e_N(p) &= J(B_N', p) - J(B_N^{+\prime}, p) \geqq 0, \\ f_N(p) &= J(A_N', p) + J(B_N^{+\prime}, p). \end{aligned}$$

If we require $A_N' \supset H_0$, then $f_N(p) \leqq 2J(H_0, p)$, and all the differences in (3.16) are defined. If as N tends to infinity

$$(3.17) \qquad N\, d_N(p)/\log N \to \infty$$

for $p \varepsilon H_0'$, then test B_N is considerably more powerful than test A_N in the sense that $P_N(B_N' \mid p)/P_N(A_N' \mid p)$ tends to zero faster than any power of N. If as N tends to infinity

$$(3.18) \qquad e_N(p)/d_N(p) \to 0$$

then the ratio of the probabilities of the error of the second kind, (3.15), tends to zero faster than any power of N for the tests B_N^+, A_N even though test B_N^+ is not necessarily more powerful than test A_N for fixed N. We summarize these results in the following theorem.

THEOREM 3.1. *Let A_N be a sequence of non-empty, non-randomized tests for testing the hypothesis $p \varepsilon H_0 \subset Q_0$ which satisfy the following conditions:*

(i) *For N sufficiently large the sets A_N, A_N' are in $B_0(p)$ for pertinent p;*

(ii) $c_N = O(\log N)$, $Nc_N/\log N \to \infty$ *as* $N \to \infty$;

(iii) *For N sufficiently large* $A_N' \supset H_0$;

then the size of test A_N for testing $p \varepsilon H_0$ is given by (3.5) and the error probability at $p \varepsilon H_0' \cap Q_0$ by (3.14). The test B_N^+ as defined in (3.9) and (3.11) satisfies, for sufficiently large N, $\alpha(B_N^+) \leqq \alpha(A_N)$ where $\alpha(A_N)$ tends to zero faster than any power of N; and for those $p \varepsilon H_0' \cap Q_0$ for which (3.17) and (3.18) hold and for which $P_N(A_N' \mid p) > 0$, the ratio of the error probabilite as given in (3.15) tends to zero faster than any power of N, as N tends to infinity.

Thus for a class of tests we have established sufficient conditions for the existence of a LRT which performs much better in the sense of Theorem 3.1 than a given test.

4. On the implementation of Theorem 3.1. In order to be able to make use of Theorem 3.1 it is necessary to investigate the function J and to explore the membership of the class $B_0(p)$ for relevant p. We shall not go into the details of the results of either of these lines of research here. However, several results of basic interest are given in this section. Some proofs, which follow by long but rather straightforward mathematical analyses, are omitted.

Even a cursory glance at Theorem 3.1 cannot fail to reveal the dominant role played by the function J and its infima. The following lemma summarizes some of the more basic properties of J.

LEMMA 4.1. *For fixed* $p \varepsilon Q_0$

(i) $J(\dot{p}, p) \geqq 0$ *with equality if and only if* $\dot{p} = p$;

(ii) J *is a convex function of* $\dot{p}$;

(iii) *For any non-empty set $B \subset Q_0$ there exists a $\dot{p}_p$ in $\bar{B} \cap Q_0$ such that $J(\dot{p}_p, p) = J(B, p)$. If $p \varepsilon \bar{B}$, the closure of B, $\dot{p}_p = p$; if not, $\dot{p}_p$ is a boundary point of B. For fixed $\dot{p} \varepsilon Q_0$*

(iv) *For any non-empty set $H_0 \subset Q_0$ there exists a point $p_{\dot{p}} \varepsilon \bar{H}_0 \cap Q_0$ such that $J(\dot{p}, p_{\dot{p}}) = J(\dot{p}, H_0)$. If $\dot{p} \varepsilon \bar{H}_0$, $p_{\dot{p}} = \dot{p}$; if not, $p_{\dot{p}}$ is a boundary point of H_0. Finally*

(v) J *is a continuous function on* $Q_0 \times Q_0$.

Because of the role of the LRT the infimum of $J(\dot{p}, p)$ subject to $J(\dot{p}, H_0) < c$ is needed for the approximation of the type II error probability in an application of Theorem 3.1. Although results exist for this case, they are not nearly as explicit as Hoeffding's [5] for the multinomial case. For example in the simple case $H_0 = \{p_0\}$ we are guaranteed a point $\dot{p}_0$ in the closure of $\{\dot{p} : J(\dot{p}, H_0) < c\}$ intersected with Q_0 such that

$$J(\dot{p}_0, p) = \inf_{\dot{p}} \{J(\dot{p}, p) : J(\dot{p}, H_0) < c\}$$

by Lemma 4.1 (iii). Use of the Lagrange multiplier technique for finding extrema yields for $p_0 = (u_0, A_0)$, $p = (u, A)$, $\dot{p}_0 = (\dot{u}_0, \dot{A}_0)$

$$\text{(4.1)} \qquad \begin{aligned} \dot{u}_0 &= u_0 + \lambda_c[\lambda_c A^{-1} + (1 - \lambda_c)A_0^{-1}]^{-1}A^{-1}(u - u_0), \\ \dot{A}_0 &= [\lambda_c A^{-1} + (1 - \lambda_c)A_0^{-1}]^{-1}, \end{aligned}$$

for a unique $0 < \lambda_c < 1$ determined by $J(\dot{p}_0, p_0) = c$. However, finding λ_c seems to be a numerical problem.

To investigate the power of A_N as compared to that of B_N^+ we have concentrated on the error probabilities. The number $d_N(p)$ is a measure of the similarity between the two tests A_N and B_N. Clearly if $A_N = B_N$, then $d_N(p) = 0$ for all p. For $A_N \neq B_N$ we wish to investigate how different the two tests should be for $N\, d_N(p)/\log N \to \infty$ with N. Or to put it another way, for what alternatives are the tests sufficiently different as measured by $N\, d_N(p)/\log N \to \infty$. For $d_N(p) > d > 0$, say, the condition (3.17) obtains and it seems likely that continuity will imply $e_N(p) \to 0$ as $N \to \infty$ so condition (3.18) will hold. We now consider the set of alternatives p, such that $d_N(p) = 0$. For convenience we drop the subscript N.

Lemma 4.2. *If $p \,\varepsilon\, \bar{B}'$, then*

(i) *$d(p) = 0$ if and only if $\{\dot{p}:J(\dot{p}, p) < J(B', p)\} \subset A$*

and

(ii) *$d(p) = 0$ implies there is $\dot{p}_p \,\varepsilon\, Q_0$ common to the boundary of A and the boundary of B such that $J(\dot{p}_p, p) = J(B', p)$.*

Proof. Only the proof of (i) will be given.

(i) If $d(p) = 0$, $J(A', p) = J(B', p)$ and

$$\{\dot{p}:J(\dot{p}, p) < J(B', p)\} = \{\dot{p}:J(\dot{p}, p) < J(A', p)\} \subset A,$$

since $\dot{p} \,\varepsilon\, A'$ implies $J(\dot{p}, p) \geqq J(A', p)$. If on the other hand $\{\dot{p}:J(\dot{p}, p) < J(B', p)\} \subset A$, then $\{\dot{p}:J(\dot{p}, p) \geqq J(B', p)\} \supset A'$ so that $J(A', p) \geqq J(B', p)$. However, $B' \subset A'$ so the reverse inequality holds, which means the equality holds and $d(p) = 0$.

The content of Lemma 4.2 is simply that, in order for $d(p) = 0$, A and B must have common boundary points and that the behavior of A near such boundary points is restricted by part (i) of the lemma.

It is generally not an easy matter to establish whether or not a subset of Q_0 is a member of $B_0(p)$. However, certain results can be stated which will help. First a result will be stated which in effect allows one to strip away extraneous parts of a subset of Q_0 and consider only that part of the subset pertinent to the probability estimate (2.38). This result also supplies the justification for the remarks following Definition 2.1.

Lemma 4.3. *Let*

$$B(c; p) = \{\dot{p}:J(\dot{p}, p) \geqq c\},\ B'(c; p) = \{\dot{p}:J(\dot{p}, p) < c\} \tag{4.2}$$

and define

$$\text{ess}_B \inf J(\dot{p}, p) = \sup \{c: m(B'(c; p) \cap B) = 0\} \tag{4.3}$$

for any measurable set $B \subset Q_0$ where m is n dimensional Lebesgue measure. Denote the $\text{ess}_B \inf J(\dot{p}, p)$ *by $c_B(p)$. Then for a measurable set B in Q_0 and point $p \,\varepsilon\, Q_0$*

(i) *$c_B(p) \geqq J(B, p)$;*

(ii) $c_{B(c;p)}(p) = J(B(c; p), p)$;
(iii) *If* $B^0 = B(c_B(p); p) \cap B$, *then* $B^0 \subset B$, $m(B^0) = m(B)$, *and*

$$J(B^0, p) = c_{B^0}(p) = c_B(p). \tag{4.4}$$

PROOF. Only the proof of (iii) will be given here.

(iii) That $B^0 \subset B$ is immediate. To show that $m(B^0) = m(B)$ it will be sufficient to show that $B'(c_B(p); p) \cap B = B^1$ has m measure zero. Let $\{c_i\}$ be a non-decreasing sequence of non-negative numbers such that $c_i \to c_B(p)$. By definition of $c_B(p)$

$$m(B'(c_i; p) \cap B) = 0, \qquad i = 1, 2, \cdots. \tag{4.5}$$

Clearly $B'(c_i; p)$ increases to $B'(c_B(p); p)$ as $i \to \infty$. Thus $B'(c_i; p) \cap B$ increases to $B'(c_B(p); p) \cap B$ and (4.5) yields

$$0 = \lim m(B'(c_i; p) \cap B) = m(B'(c_B(p); p) \cap B) \tag{4.6}$$

so that $m(B^1) = 0$.

Now $c_{B^0}(p) = \sup\{c : m(B'(c; p) \cap B^0) = 0\}$ and $B'(c; p) \cap B^0$ is the empty set for $c < c_B(p)$ so $c_{B^0}(p) \geqq c_B(p)$. If $c > c_B(p)$, $m(B'(c; p) \cap B^0) = m((B'(c; p) \cap B^0) \cup (B'(c; p) \cap B^1)) = m(B'(c; p) \cap B) > 0$ by definition of $c_B(p)$ and the fact that $m(B^1) = 0$. Thus $c_{B^0}(p) \leqq c_B(p)$ and it follows that $c_{B^0}(p) = c_B(p)$.

Suppose $J(B^0, p) < c_{B^0}(p)$; then there is a number c_0 such that

$$J(B^0, p) < c_0 < c_{B^0}(p) \tag{4.7}$$

and $\{\dot{p} : J(\dot{p}, p) \leqq c_0\} \cap B^0$ is empty by definition of B^0. Thus

$$\inf\{J(\dot{p}, p) : \dot{p} \in B^0\} = \inf\{J(\dot{p}, p) : \dot{p} \in \{\dot{p} : J(\dot{p}, p) > c_0\} \cap B^0\} \tag{4.8}$$

and thus $J(B^0, p) \geqq c_0$ which is a contradiction. Thus we have $J(B^0, p) = c_{B^0}(p)$, since by (i) $J(B^0, p) \leqq c_{B^0}(p)$. This concludes the proof.

Because of the possible statistical applications of the probability estimate (2.38), e.g. Theorem 3.1, it is of particular interest to be able to decide whether or not sets of the form $\{\dot{p} : F(\dot{p}) \geqq 0\}$ for suitably restricted F belong to $B_0(p)$. The following result provides sufficient conditions for deciding this question in the affirmative.

LEMMA 4.4. *Let* w *be a positive constant and define* $s(w)$ *by*

$$s(w) = \inf\{\|p_1 - p_2\| : F(p_1) = 0, F(p_2) = w\}$$

and B_w *by*

$$B_w = \{\dot{p} : F(\dot{p}) \geqq w\}.$$

Suppose F *is continuous in* Q_0 *and* $\{\dot{p} : F(\dot{p}) \geqq 0\} \subset Q_0$. *If for* $p \in Q_\epsilon \cap \{\dot{p} : F(\dot{p}) < 0\}$ *there exists a* $w > 0$ *such that*

(i) $s(w) > \delta'$, *for* δ' *as defined in* (2.30),

(ii) $c_{B_w}(p) \leqq c_B(p) + \frac{3}{2}\delta$, *for* $B = \{\dot{p}:F(\dot{p}) \geqq 0\}$
where δ *is sufficiently small so that*

$$\{\dot{p}:J(\dot{p}, p) = c_B(p) + \tfrac{3}{2}\delta\} \cap \{\dot{p}:F(\dot{p}) = 0\} \neq \varnothing,$$

then $\{\dot{p}:F(\dot{p}) \geqq 0\} \, \varepsilon \, B_0(p)$.

5. An example for $k = 2$. In this example we wish to test the composite hypothesis $p = (u, A_0)$ for fixed A_0, where $(u, A_0) \, \varepsilon \, Q_0$ for all u. By a simple change of variables we can, without loss of generality, consider the case $A_0 = I$, the k by k identity matrix. Thus we wish to test the hypothesis $p \, \varepsilon \, H_0$

$$H_0 = \{p:p = (u, A), A = I\} \tag{5.1}$$

versus $p \not\varepsilon H_0$.

For this hypothesis we compare a test developed by S. N. Roy, [6], Section 6.4, with the LRT for this hypothesis. Our aim will be to show that the ratio of the error probability of LRT to that of Roy's test tends to zero faster than any power of N as $N \rightarrow \infty$ for most alternatives, provided the size of the tests tend to zero sufficiently fast. The LRT for this hypothesis as given by (3.2) rejects when $J(p, H_0)$ exceeds a constant. For the set H_0 under consideration

$$J(\dot{p}, H_0) = \tfrac{1}{2}[\operatorname{tr} \dot{A} - k - \log |\dot{A}|]. \tag{5.2}$$

Roy's test for $p \, \varepsilon \, H_0$ vessus $p \not\varepsilon H_0$ is developed under his heuristic union-intersection principle and rejects for $\dot{p} \, \varepsilon \, R$ where

$$R' = \{p:c_1 < \lambda_1(\dot{A}), \lambda_k(\dot{A}) < c_2\} \tag{5.3}$$

for positive constants $0 < c_1 < c_2$. Here the $\lambda_i(\dot{A})$, $i = 1, \cdots, k$, are again the ordered eigenvalue of $\dot{A}$. The constants c_1 and c_2 are determined by the required size of the test and the requirement of local unbiasedness, i.e., the power function should have a local minimum at each point of the boundary of R'. We shall not be concerned with the choice of c_1 and c_2 here and shall only require that $c_1 < 1 < c_2$ so that $H_0 \subset R'$. We shall assume c_1 and c_2 are independent of N.

In order to be able to apply Theorem 3.1 to this example we must verify that certain probability estimates hold. Thus we should like to know that for $p \, \varepsilon \, R' \cap Q_\epsilon$, $R \, \varepsilon \, B_0(p)$; and for $p \, \varepsilon \, R \cap Q_\epsilon$, $R' \, \varepsilon \, B_0(p)$. It is relatively simple to show the former holds. This result follows from the following lemma.

Lemma 5.1. *As defined in* (5.3), R' *satisfies* (i) $R' \subset Q_0$ (ii) R' *is convex.*

We omit the proof.

Because of the difficulty involved in showing $R' \, \varepsilon \, B_0(p)$ for $p \, \varepsilon \, R \cap Q_\epsilon$ we are forced to consider the case $k = 2$. In this case a long, tedious, elementary argument employing Lemma 4.4 establishes the following theorem.

Theorem 5.1. *For* $p \, \varepsilon \, R \cap Q_\epsilon$, $R' \, \varepsilon \, B_0(p)$.

We now consider those alternatives to the hypothesis H_0 for which $d(p) = 0$ in the case $k = 2$, i.e.,

$$\{p:d(p) = J(B', p) - J(R', p) = 0\} \tag{5.4}$$

where R' is given by (5.3) for $k = 2$ and

$$(5.5)\quad B' = \{\dot{p}: J(\dot{p}, H_0) < J(R, H_0) = \min\{c_1 - 1 - \log c_1, c_2 - 1 - \log c_2\}\}.$$

In view of (5.2), we shall focus our attention on the space of eigenvalues of $\dot{A}$, i.e.

$$(5.6)\qquad \{(\lambda_1, \lambda_2): 0 < \lambda_1 \leqq \lambda_2\}.$$

A straight forward analysis establishes the following theorem.

THEOREM 5.2. *If R' and B' are as defined in* (5.3) *and* (5.5) *respectively, then*

$$(5.7)\quad \begin{aligned}\{p: d(p) = 0\} &= \{p: \lambda_1(A) < c_1, \lambda_2(A) = 1\} \\ &\qquad \cup \{p: \lambda_1(A) = 1, \lambda_2(A) > c_2\}, \quad \text{I.} \\ &= \{p: \lambda_1(A) < c_1, \lambda_2(A) = 1\}, \quad \text{II.} \\ &= \{p: \lambda_1(A) = 1, \lambda_2(A) > c_2\}, \quad \text{III.}\end{aligned}$$

and in any case

$$\{p: d(p) = 0\} \subset \{p: \lambda_1(A) < c_1, \lambda_2(A) = 1\} \cup \{p: \lambda_1(A) = 1, \lambda_2(A) > c_2\}.$$

Here I, II, III *stand for the conditions that $c_1 - 1 - \log c_1$ is equal to, less than, or greater than $c_2 - 1 - \log c_2$ respectively.*

COROLLARY 5.1. *If the hypotheses of Theorem* 5.2 *are satisfied and $p \varepsilon R$ is such that $\lambda_1(A) \neq 1$ and $\lambda_2(A) \neq 1$, then $d(p) > 0$.*

We now wish to consider

$$(5.8)\qquad e_N(p) = J(B', p) - J(B_N^{+\prime}, p)$$

for

$$(5.9)\qquad B_N^{+\prime} = \{\dot{p}: J(\dot{p}, H_0) \leqq J(R, H_0) + \Delta_N\}$$

where $\Delta_N \to 0$ as $N \to \infty$. Now if $B_c' = \{\dot{p}: J(\dot{p}, H_0) \leqq c\}$, then to find $J(B_c', p)$ we consider the problem of minimizing $J(\dot{p}, p)$ subject to $J(\dot{p}, H_0) = c$, and note that we are guaranteed a point $\dot{p}_c$ in boundary of B_c' for $p \varepsilon B_c$ so that $J(\dot{p}_c, p) = J(B_c', p)$. By the LaGrange multiplier criterion and the regularity of $J(\dot{p}, H_0)$ in this case, p_c must satisfy

$$(5.10)\qquad \nabla(J(\dot{p}, p) - \mu J(\dot{p}, H_0)) = 0,\ J(\dot{p}, H_0) = c.$$

The implicit function theorem guarantees that for the H_0 under consideration here, $\dot{p}_c$ is a continuous function of c. Thus $J(B_c', p) = J(\dot{p}_c, p)$ is a continuous function of c and it follows that $e_N(p) \to 0$ as $N \to \infty$.

Thus Theorem 3.1 applies for those $p \varepsilon R \cap Q_\epsilon$ for which $\lambda_1(A) \neq 1$ and $\lambda_2(A) \neq 1$ and thus for these p we have that the ratio of the type II error probability of a LRT for testing $p \varepsilon H_0$ vs. $p \not\varepsilon H_0$ to that of Roy's test for the same hypothesis tends to zero faster than any power of N as $N \to \infty$ in the case $k = 2$.

REFERENCES

[1] Bahadur, R. R. and Rao, R. Ranga (1960). On deviations of the sample mean. *Ann. Math. Statist.* **31** 1015–1027.

[2] Borovkov, A. A. and Rogozin, B. A. (1965). On the central limit theorem in the multidimensional case (in Russian). *Teor. Verojatnost. i Primenen.* **10** 61–69.

[3] Chernoff, H. (1952). A measure of asymptotic efficiency for tests of a hypothesis based on the sum of observations. *Ann. Math. Statist.* **23** 493–507.

[4] Hoeffding, Wassily (1965). *On Probabilities of Large Deviations.* Institute of Statistics Mimeo Series No. 443, Chapel Hill, North Carolina.

[5] Hoeffding, Wassily (1965). Asymptotically optimal tests for multinomial distributions. *Ann. Math. Statist.* **36** 369–401.

[6] Roy, S. N. (1957). *Some Aspects of Multivariate Analysis.* Wiley, New York.

[7] Sanov, I. N. (1965). On the probability of large deviations of random variables. *Selected Translations of Mathematical Statistics and Probability* (English Translation). **1** 213–244.

[8] Wald, A. (1943). Tests of statistical hypotheses concerning several parameters when the number of observations is large. *Trans. Am. Math. Soc.* **54** 426–482.