ASYMPTOTICALLY OPTIMAL TESTS FOR MULTINOMIAL DISTRIBUTIONS

Wassily Hoeffding

Institute of Statistics
Mimeograph Series No. 396
July 1964
ASYMPTOTICALLY OPTIMAL TESTS FOR MULTINOMIAL DISTRIBUTIONS

by

Wassily Hoeffding

July 1964

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

Institute of Statistics
Mimeo Series No. 396
Corrections to
"Asymptotically optimal tests for multinomial distributions"
by
Wassily Hoeffding
(University of North Carolina, Institute of Statistics
Mimeo Series No. 396)

On p. 5.4 delete the line following (5.18).

On p. 5.5 replace the three lines following (5.22) by: The following lemma will be
used in the next section.

On p. 5.5 delete the three-line sentence preceding (5.24).

On pp. 7.1-7.9 the letter ε is blurred in many places.

<table>
<thead>
<tr>
<th>Page</th>
<th>Line</th>
<th>As it appears</th>
<th>Correction</th>
</tr>
</thead>
<tbody>
<tr>
<td>P. 4.2, Lemma 4.2</td>
<td>y ∈ $\Omega$</td>
<td>y ≠ $\Omega$</td>
<td></td>
</tr>
<tr>
<td>P. 4.4, Lemma 4.5(b)</td>
<td>$\sum p_i = 0$</td>
<td>$p_i ≠ 0$</td>
<td></td>
</tr>
<tr>
<td>P. 5.3, line 5</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>P. 6.1, paragraph after Definition 6.1</td>
<td>relative to $p$ or $\Lambda$</td>
<td>relative to $p$</td>
<td></td>
</tr>
<tr>
<td>P. 6.3, Theorem 6.3(II)</td>
<td>point of $y$</td>
<td>point $y$</td>
<td></td>
</tr>
<tr>
<td>P. 6.4, line 6</td>
<td>$p(y)$</td>
<td>$\hat{p}(y)$</td>
<td></td>
</tr>
<tr>
<td>P. 6.4, line 9</td>
<td>$p_i(y)$</td>
<td>$\hat{p}_i(y)$</td>
<td></td>
</tr>
<tr>
<td>P. 7.3, line 3 from bottom</td>
<td>for any $p(x)$</td>
<td>for any $\epsilon$</td>
<td></td>
</tr>
<tr>
<td>P. 8.1, line 10 from bottom</td>
<td>$p(z(N))$</td>
<td>$\hat{p}(z(N))$</td>
<td></td>
</tr>
<tr>
<td>P. 8.2, line 4</td>
<td>$p_{ij}(x)$</td>
<td>$\hat{p}_{ij}(x)$</td>
<td></td>
</tr>
<tr>
<td>P. 8.2, line after (8.3)</td>
<td>$\sum_{i=1}^{k} (x_i - \bar{p}_i)^2$</td>
<td>$\sum_{i=1}^{k} (\hat{x}_i - \hat{p}_i)^2$</td>
<td></td>
</tr>
<tr>
<td>P. 9.1, (9.1)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>P. A1, line 10</td>
<td>replaced</td>
<td>replaced by</td>
<td></td>
</tr>
<tr>
<td>P. A3, line 8 from bottom</td>
<td>similar</td>
<td>similar to</td>
<td></td>
</tr>
<tr>
<td>P. A5, line 7 from bottom</td>
<td>$(y_i^{(N)} - \hat{y}_i^{(N)})$</td>
<td>$(Z_i^{(N)} - \hat{y}_i^{(N)})$</td>
<td></td>
</tr>
</tbody>
</table>
Summary. Tests of simple and composite hypotheses for multinomial distributions are considered. It is assumed that the size $\frac{\alpha}{N}$ of a test tends to 0 at a suitably rapid rate as the sample size $N$ increases. The main concern of this paper is to substantiate the following proposition: If a given test of size $\frac{\alpha}{N}$ is "sufficiently different" from a likelihood ratio test then there is a likelihood ratio test of size $\leq \frac{\alpha}{N}$ which is considerably more powerful than the given test at "most" points in the set of alternatives when $N$ is large enough. In particular, it is shown that chi-square tests of simple and of some composite hypotheses are inferior, in the sense described, to the corresponding likelihood ratio tests. Certain Bayes tests are shown to share the above-mentioned property of a likelihood ratio test.

---

1This research was supported by the Mathematics Division of the Air Force Office of Scientific Research.
1. Introduction. This paper is concerned with asymptotic properties of tests of simple and composite hypotheses concerning the parameter vector \( p = (p_1, \ldots, p_k) \) of a multinomial distribution as the sample size \( N \) tends to infinity. In traditional asymptotic test theory the size \( \alpha \) of the test is held fixed and its power is investigated at alternatives \( p = p(N) \) which approach the hypothesis set as \( N \to \infty \), in such a way as to keep the error probability away from 0. These restrictions make it possible to apply the central limit theorem and its extensions. However, it seems reasonable to let the size \( \alpha_N \) of a test tend to 0 as the number \( N \) of observations increases. It is also of interest to consider alternatives not very close to the hypothesis, at which, typically, the error probabilities will tend to zero. To attack these problems, the theory of probabilities of large deviations is needed. For the case of sums of independent random variables this theory is by now well developed. It has been used by Chernoff [1] to compare the performance of tests based on sums of independent, identically distributed random variables when the error probabilities tend to zero. Sanov [2] made an interesting contribution toward a general theory of probabilities of large deviations. He studied the asymptotic behavior of the probability that the empirical distribution function is contained in a given set \( A \) of distribution functions when the true distribution function is not in \( A \). For the special case of a multinomial distribution a slight elaboration of one of Sanov's results implies the following.
Let the random vector \( Z(N) \) take the values \( Z(N) = (n_1/N, \ldots, n_k/N) \), where \( n_1, \ldots, n_k \) are nonnegative integers whose sum is \( N \), and let the probability of \( Z(N) = z \) be \( N! \prod_{i=1}^{k} (p_i^{n_i}/n_i!) \), where \( p = (p_1, \ldots, p_k) \in \Omega \), the set of points \( p \) with \( p_1 \geq 0, p_1 + \cdots + p_k = 1 \). Let \( A \) be any subset of \( \Omega \), and let \( A(N) \) denote the set of points \( z(N) \) which are in \( A \). Then for the probability \( P_N(A|p) \) of \( z(N) \subseteq A \) we have (see theorem 2.1)

\[
(1.1) \quad P_N(A|p) = \exp \left\{ -N I(A(N), p) + O(\log N) \right\},
\]

uniformly for \( A \subseteq \Omega \) and \( p \in \Omega \), where

\[
(1.2) \quad I(x, p) = \sum_{i=1}^{k} x_i \log \left( \frac{x_i}{p_i} \right),
\]

\[
(1.3) \quad I(A, p) = \inf \{ I(x, p) \mid x \in A \}.
\]

This elementary and crude estimate of the probability \( P_N(A|p) \) makes it possible to study, to a first approximation, the asymptotic behavior of the error probabilities of an arbitrary (non-randomized) test of a hypothesis concerning \( p \) when these probabilities tend to 0 at a sufficiently rapid rate.

In section 3 the special role of the likelihood ratio test is brought out. Let \( H \) be the hypothesis that \( p \in \Lambda \) (\( \Lambda \subseteq \Omega \)). The likelihood ratio test, based on an observation \( z(N) \) of \( Z(N) \), for testing \( H \) against the alternatives \( p \in \Omega - \Lambda \) rejects \( H \) when

\[
I(z(N), \Lambda) > \text{const},
\]

where \( I(x, \Lambda) = \inf \{ I(x, p) \mid p \in \Lambda \} \). For the size of an arbitrary test which rejects \( H \) when \( z(N) \in A \) we have from (1.1)

\[
(1.4) \quad \sup_{p \in \Lambda} P_N(A|p) = \exp \left\{ -N I(A(N), \Lambda) + O(\log N) \right\},
\]

uniformly for \( A \subseteq \Omega \) and \( \Lambda \subseteq \Omega \), where \( I(A, \Lambda) = \inf \{ I(x, p), x \in A, p \in \Lambda \} \).
This easily implies the following: The union of the critical regions \( A^{(N)} \) of all tests of size \( \leq \alpha_N \) for testing \( H \) is contained in the critical region \( B^{(N)} \) of a likelihood ratio test for testing \( H \) against \( p \in \Omega - \Lambda \) whose \( \alpha_N \) satisfies

\[
\log \alpha_N = \log \alpha_N + O(\log N).
\]

Thus if \( \alpha_N \) tends to 0 faster than any power of \( N \), the size of the \( B^{(N)} \) test is, to a first approximation, \( \alpha_N \). Of course, \( \alpha_N' \geq \alpha_N \). It is trivial that the \( B^{(N)} \) test is uniformly at least as powerful as any test of size \( \leq \alpha_N \).

We can also define a likelihood ratio test of size \( \leq \alpha_N \) whose critical region does not differ much from \( B^{(N)} \) in the sense that both critical regions are of the form

\[
N I(z^{(N)}, \Lambda) \geq - \log \alpha_N + O(\log N).
\]

The main concern of this paper is to substantiate the following proposition:

If a given test of size \( \alpha_N \) is "sufficiently different" from a likelihood ratio test, then there is a likelihood ratio test of size \( \leq \alpha_N \) which is considerably more powerful than the given test at "most" points \( p \) in the set of alternatives when \( N \) is large enough, provided that \( \alpha_N \to 0 \) at a suitable rate. The meaning of the words in quotation marks will have to be made precise. By "considerably more powerful" we mean that the ratio of the error probabilities at \( p \) of the two tests tends to 0 more rapidly than any power of \( N \).

A general characterization of the set \( \Gamma_N \) of alternatives \( p \) at which a given test is considerably less powerful than a comparable likelihood ratio test is contained in theorem 3.1. Sections 4 and 5 are preparatory to what follows and deal with properties of the function \( I(x, p) \) and its infima. In section 6 we restrict ourselves to tests whose critical regions are regular in a sense which implies that the expression (1.4) for the size of a test remains true with
I(A(N), Λ) replaced by I(A, Λ), the infimum of I(x, Λ) with respect to all x ∈ A (not only with respect to the lattice points z(N) contained in A), and an analogous replacement may be made in the expressions for the error probabilities $P_N(A' \mid p)$, where $A' = Ω - A$ and $p \in Ω - Λ$. (Sufficient conditions for regularity are given in the Appendix.) Consider a test which rejects H when $z(N) \in A$ (where $A = A_N$ may depend on N). Let $B = \{ x \mid I(x, Λ) \geq I(A, Λ) \}$, so that $B(N)$ is the critical region of a likelihood ratio test. Note that $A \subseteq B$.

It is shown that the set $Γ_N$ essentially depends on the set of common boundary points of the sets A and B. In particular, if the A test differs sufficiently from a likelihood ratio test in the sense that the sets A and B have only finitely many boundary points in common then, under certain additional conditions, a likelihood ratio test whose size does not exceed the size of the A test is considerably more powerful than the latter at all alternatives except those points $p$ which lie on certain curves in the $(k-1)$-dimensional simplex $Ω$ and those at which both tests have zero error probabilities.

In section 7 the result just described is shown to be true for a chi-square test of a simple hypothesis whose size tends to 0 at a suitable rate (theorem 7.4). This is of special interest in view of the fact that if the size of the chi-square test tends to a positive limit, its critical region and power differ little from those of a likelihood ratio test. In section 8 chi-square tests of composite hypotheses are briefly discussed. An example shows that at least in some cases the situation is similar to that in the case of a simple hypothesis. It is noted that one common version of the chi-square test may have the property that its size cannot be smaller than some power of N, which makes the theory of this paper inapplicable. Certain competitors of the chi-square test are considered in section 9.
It is pointed out that certain Bayes tests have the same asymptotic power properties as the corresponding likelihood ratio test (section 10).

The likelihood ratio test was introduced by J. Neyman and E. S. Pearson in 1928 [3]. It is known that the likelihood ratio test has certain asymptotically optimal properties when the error probabilities are bounded away from 0 (Wald [4]). The present results are of a different nature and appear to be of a novel type.

An extension of the results of this paper to certain classes of distributions other than the multinomial class should be possible. (The extension to the case of several independent multinomial random vectors is quite straightforward.)
2. Probabilities of large deviations in multinomial distributions: Crude estimates. Let \( Z^{(N)} = (Z_1^{(N)}, \ldots, Z_k^{(N)}) \) be a random vector whose values are

\[
Z^{(N)} = (Z_1^{(N)}, \ldots, Z_k^{(N)}) = \left( \frac{n_1}{N}, \ldots, \frac{n_k}{N} \right),
\]

where \( n_1, \ldots, n_k \) are any nonnegative integers such that \( n_1 + \ldots + n_k = N \), and whose distribution is given by

\[
\Pr\{ Z^{(N)} = z^{(N)} \} = P_N(z^{(N)} | p) = \frac{N!}{n_1! \cdots n_k!} p_1^{n_1} \cdots p_k^{n_k}.
\]

Here \( p = (p_1, \ldots, p_k) \) is any point in the simplex

\[
\Omega = \{ (x_1, \ldots, x_k) \mid x_1 \geq 0, \ldots, x_k \geq 0, x_1 + \ldots + x_k = 1 \}
\]

By convention, \( p_i = 1 \) if \( p_i = n_i = 0 \).

We can write

\[
P_N(z^{(N)} | p) = P_N(z^{(N)} | z^{(N)}) \exp \{ -N I(z^{(N)}, p) \},
\]

where, for any two points \( x \) and \( p \) in \( \Omega \),

\[
I(x, p) = \sum_{i=1}^{k} x_i \log \left( \frac{x_i}{p_i} \right).
\]

Here it is understood that \( x_i \log \left( \frac{x_i}{p_i} \right) = 0 \) if \( x_i = 0 \).

We note that \( I(x, p) > 0 \) unless \( x = p \) (since \( \log u > 1 - u^{-1} \) for \( u > 0, u \neq 1 \)). Also, \( I(x, p) < \infty \) unless \( p_i = 0 \) and \( x_i > 0 \) for some \( i \).

For any subset \( A \) of \( \Omega \) let

\[
P_N(A | p) = \Pr \{ Z^{(N)} \in A \} = \sum_{z^{(N)} \in A} P_N(z^{(N)} | p).
\]

The set of lattice points \( z^{(N)} \) contained in \( A \) will be denoted by \( A^{(N)} \). We define

\[
I(A, p) = \inf \{ I(x, p) \mid x \in A \}, \quad I(A, p) = +\infty \text{ if } A \text{ is empty}.
\]
The following theorem is a slight elaboration of a result due to Sanov [2].

**Theorem 2.1.** For any set \( A \subset \Omega \) and any point \( p \in \Omega \) we have

\[
(2.8) \quad C_0 N^{-\frac{(k-1)}{2}} \exp \left\{-N I(A^{(N)}, p)\right\} \leq P_N(A \mid p) \leq (N^{k-1}) \exp \left\{-N I(A^{(N)}, p)\right\},
\]

where \( C_0 \) is a positive absolute constant. Hence

\[
(2.9) \quad P_N(A \mid p) = \exp \left\{-N I(A^{(N)}, p) + O\left(\log N\right)\right\},
\]

uniformly for \( A \subset \Omega \) and \( p \in \Omega \). Also,

\[
(2.10) \quad P_N(A \mid p) \leq \exp \left\{-N I(A, p) + O\left(\log N\right)\right\},
\]

uniformly for \( A \subset \Omega \) and \( p \in \Omega \).

**Proof.** Clearly (2.8) implies (2.9) and since \( A^{(N)} \subset A \), (2.8) implies (2.10). It is sufficient to prove (2.8).

If \( A^{(N)} \) is empty, \( P_N(A \mid p) = 0 \) and (2.8) is trivially true. Assume that \( A^{(N)} \) is not empty.

The number of points \( z^{(N)} \) in \( \Omega \) is easily found to be \( \binom{N+k-1}{k-1} \). By (2.4), \( z^{(N)} \in A \) implies \( P_N(z^{(N)} \mid p) \leq \exp \left\{-N I(A^{(N)}, p)\right\} \). Hence the second inequality (2.8) follows from (2.6).

By Stirling's formula, for \( m \geq 1 \),

\[
m! = m^m (2\pi m)^{\frac{1}{2}} \exp \left(-m + \frac{\theta}{12m}\right), \quad 0 < \theta < 1.
\]

Hence it easily follows that if \( n_i \geq 1 \) for all \( i \),

\[
(2.11) \quad P_N(z^{(N)} \mid z^{(N)}) = \frac{N!}{N^n} \prod_{i=1}^{k} \frac{n_i!}{n_i^i} \geq C_0 N^{-\frac{k-1}{2}},
\]

where \( C_0 \) is a positive absolute constant. (We can take \( C_0 = \frac{1}{2} \).) If \( n_i = 0 \) for some \( i \), (2.11) is a fortiori true. The first inequality (2.8) follows from (2.4), (2.6) and (2.11).
Theorem 2.1 is nontrivial only if the set $A$ contains no points $z^{(N)}$ which are too close to $p$. In this sense the theorem is concerned with probabilities of large deviations of $z^{(N)}$ from its mean $p$.

It should be noted that (2.9) gives an asymptotic expression for the logarithm of the probability on the left but not for the probability itself. This crude result is sufficient to study asymptotically the main features of any test whose size tends to 0 fast enough as $N$ increases.
3. The role of the likelihood ratio test. Consider the problem of testing, on the basis of an observation $z^{(N)}$ of the random vector $Z^{(N)}$, the hypothesis $H$ that the parameter vector $p$ is contained in a subset $\Lambda$ of $\Omega$. The likelihood ratio test for testing $H$ against the alternative $p \in \Lambda' = \Omega - \Lambda$ is based on the statistic

$$
\sup \frac{\sup \{ P_N(z^{(N)} | p) : p \in \Lambda \}}{\sup \{ P_N(z^{(N)} | p) : p \in \Omega \}} = \exp \{-N I(z^{(N)}, \Lambda)\},
$$

where

$$
I(x, \Lambda) = \inf \{ I(x, p) : p \in \Lambda \}.
$$

The equality in (3.1) follows from (2.4). Thus the likelihood ratio test rejects $H$ if $I(z^{(N)}, \Lambda)$ exceeds a constant.

Now consider an arbitrary test which rejects $H$ is $z^{(N)} \in \Lambda_N$, where $\Lambda_N$ is any subset of $\Omega$. For the size of the test (the supremum of its error probability for $p \in \Lambda$) we have from theorem 2.1

$$
\sup \{ P_N(\Lambda_N | p) : p \in \Lambda \} = \exp \{-N I(\Lambda_N^{(N)}, \Lambda) + O(\log N)\},
$$

uniformly for $\Lambda_N \subset \Omega$ and $\Lambda \subset \Omega$, where

$$
I(\Lambda, \Lambda) = \inf \{ I(x, p) : x \in \Lambda, p \in \Lambda \}.
$$

Clearly $I(\Lambda, \Lambda) = \inf \{ I(x, \Lambda) : x \in \Lambda \} = \inf \{ I(\Lambda, p), p \in \Lambda \}$.

The test of the preceding paragraph will be referred to as test $\Lambda_N$. The set $\Lambda_N^{(N)}$ of all points $z^{(N)}$ contained in $\Lambda_N$ will be called its critical region.

(We could have assumed that $\Lambda_N$ contains no other points than the lattice points $z^{(N)}$, but it is often convenient to define the critical region in terms of a more inclusive set.)

We may compare the test $\Lambda_N$ with the likelihood ratio test which rejects $H$ if $z^{(N)} \in B_N$, where
(3.5) \[ B_N = \{ x \mid I(x, \Lambda) \geq c_N \} \quad c_N = I(A_N^{(N)}, \Lambda) \].

Its critical region \( B_N^{(N)} \) contains the critical region \( A_N^{(N)} \). In fact, \( B_N^{(N)} \) is the union of the critical regions of all tests \( A_N^{(N)} \) for which \( I(A_N^{(N)}, \Lambda) \geq c_N \).

Moreover, the size of the test \( B_N \) is \( \exp \{-N^{-1} c_N + O(\log N)\} \), since \( I(B_N^{(N)}, \Lambda) = c_N \). Thus if \( Nc_N/\log N \) tends to infinity with \( N \), which means that the size \( \alpha_N \) of the test \( A_N \) tends to 0 faster than any power of \( N \), then the size \( \alpha_N^* \) of the test \( B_N \) is approximately equal to \( \alpha_N \) in the sense that \( \log \alpha_N^* = \log \alpha_N + O(\log N) \).

In a similar way we can obtain the following conclusion: The union of the critical regions of all tests of size \( \leq \alpha_N \) for testing the hypothesis \( p \in \Lambda \) is contained in the critical region of a likelihood ratio test for testing \( p \in \Lambda \) against \( p \notin \Lambda \) whose size \( \alpha_N^\ast \) satisfies \( \log \alpha_N^\ast = \log \alpha_N + O(\log N) \). The simple proof is left to the reader.

Since \( A_N^{(N)} \subset B_N^{(N)} \), the probability that test \( B_N \) rejects \( H \) is never smaller than the probability that test \( A_N \) rejects \( H \). Thus \( B_N \) is uniformly at least as powerful as \( A_N \), but the size of \( B_N \) is in general somewhat larger than the size of \( A_N \). It may be more appropriate to compare a given test with a likelihood ratio test whose size does not exceed the size of the former. Now it easily follows from (3.3) that we can choose numbers

(3.6) \[ 0 \leq \delta_N = O(N^{-1} \log N) \]

such that the size of the likelihood ratio test

(3.7) \[ B_N^* = \{ x \mid I(x, \Lambda) \geq c_N + \delta_N \} \]

is not larger than the size of test \( A_N \). If \( Nc_N/\log N \) tends to infinity fast enough, we may expect that the power of the test \( B_N^* \) will not be much smaller than the power of \( B_N \).
For any \( p \in \Lambda' (\Lambda' = \Omega - \Lambda) \) the probabilities that the tests \( A_N \) and \( B_N \) falsely accept the hypothesis are given by

\[
P_N(A_N^+ | p) = \exp \left\{ -N I(A_N^+(N), p) + o(\log N) \right\},
\]

\[
P_N(B_N^+ | p) = \exp \left\{ -N I(B_N^+(N), p) + o(\log N) \right\}.
\]

Always \( P_N(B_N^+ | p) \leq P_N(A_N^+ | p) \) and \( I(B_N^+(N), p) \geq I(A_N^+(N), p) \). At those points \( p \in \Lambda' \) for which \( P_N(A_N^+ | p) \neq 0 \) and

\[
\lim_{N \to \infty} \frac{N\{I(B_N^+(N), p) - I(A_N^+(N), p)\}}{\log N} = + \infty,
\]

the test \( B_N \) is considerably more powerful than \( A_N \) in the sense that the ratio of the error probabilities at \( p \), \( P_N(B_N^+ | p)/P_N(A_N^+ | p) \), tends to 0 more rapidly than any power of \( N \).

For the test \( B_N^* \) whose size does not exceed the size of \( A_N \) we have a similar conclusion. Note that \( B_N^* \) is not necessarily more powerful than \( A_N \), and the difference in (3.10) with \( B_N \) replaced by \( B_N^* \) may be negative. However, if \( P_N(A_N^+ | p) \neq 0 \), (3.10) is satisfied, and

\[
\lim_{N \to \infty} \frac{I(B_N^+(N), p) - I(B_N^*(N), p)}{I(B_N^+(N), p) - I(A_N^+(N), p)} = 0
\]

then the ratio \( P_N(B_N^* | p)/P_N(A_N^+ | p) \) tends to 0 more rapidly than any power of \( N \).

The main conclusions of the preceding discussion are summarized in the following theorem.

**Theorem 3.1.** Let \( \Lambda \) and \( A_N \) be non-empty subsets of \( \Omega \). Then

(a) the size of the test \( A_N \) for testing the hypothesis \( p \in \Lambda \) is given by (3.3) and its error probability at \( p \in \Lambda' \) by (3.8).

(b) There exist positive numbers \( \delta_N = O(N^{-1} \log N) \) such that the size of the likelihood ratio test \( B_N^* = \{ x \mid I(x, \Lambda) \geq I(A_N^+(N), \Lambda) + \delta_N \} \) does not exceed the size of the test \( A_N \).
(c) For each \( p \in \Lambda' \) such that \( P_N(A'_N | p) \neq 0 \) and conditions (3.10) and (3.11) are satisfied, the ratio \( P_N(B'_N | p)/P_N(A'_N | p) \) of the error probabilities at \( p \) tends to 0 faster than any power of \( N \).

In section 6 we shall continue in more detail the study of the set of alternatives at which a given test is less powerful than a comparable likelihood ratio test, assuming that the sets \( A_N \) are regular in a certain sense. The following two sections are preparatory to what follows.
4. The function $I(x, p)$ and its infima. In this section properties of the function

$$I(x, p) = \sum_{i=1}^{k} x_i \log \left( \frac{x_i}{p_i} \right)$$

and its infima for $x \in A$ or $p \in \Lambda$ are studied.

The function $I(x, p)$ is defined for $x$ and $p$ in the simplex $\Omega$ given by (2.3). Throughout, $\Omega$ is considered as the space of the points $x$ and $p$, with the Euclidean metric. Thus the complement $A'$ of a subset $A$ of $\Omega$ is $\Omega - A$. The closure of $A$ is denoted by $\overline{A}$. The boundary of $A$ is $\overline{A} \cap \overline{A}'$.

We define the subsets $\Omega_o$ and $\Omega(p)$ (for $p \in \Omega$) of $\Omega$ by

(4.1) $\Omega_o = \{ x \mid x_i > 0, \ i = 1, \ldots, k \}$,

(4.2) $\Omega(p) = \{ x \mid x_i = 0 \text{ if } p_i = 0 \}$.

Thus if $p \in \Omega_o$, $\Omega(p) = \Omega$. If $p \in \Omega_o'$, $\Omega(p)$ is the intersection of those faces $\{ x \mid x_i = 0 \}$ of the simplex $\Omega$ for which $p_i = 0$.

**Lemma 4.1.** (a) $0 \leq I(x, p) \leq \infty$. $I(x, p) = 0$ if and only if $x = p$.

(b) If $p \in \Omega_o$, $I(\cdot, p)$ is continuous and bounded in $\Omega$. For each $p \in \Omega_o'$, $I(\cdot, p)$ is continuous and bounded in $\Omega(p)$.

(c) For each $x \in \Omega$, $I(x, \cdot)$ is continuous in $\Omega$. That is, $p^j \to p$ implies $I(x, p^j) \to I(x, p)$, even when $I(x, p) = \infty$.

(d) For each $p \in \Omega$, $I(\cdot, p)$ is convex in $\Omega$. For each $x \in \Omega$, $I(x, \cdot)$ is convex in $\Omega$.

**Proof.** (a) See section 2 after (2.5).

(b) If $p \in \Omega_o$, $I(\cdot, p)$ is bounded since $I(x, p) \leq \Sigma x_i \log \left( \frac{1}{p_i} \right) \leq \max_i \log \left( \frac{1}{p_i} \right)$. The proof of continuity is obvious. For $p \in \Omega_o'$ the proof is similar.
(c) If \( x \in \Omega(p) \) then \( I(x, p) < \infty \) and the continuity at \( p \) is obvious. If \( x \in \Omega'(p) \) then \( I(x, p) = \infty \). If \( p' \to p \) then \( p_i' \to 0 \) for some \( i \) with \( x_i > 0 \). Hence \( I(x, p') \to \infty = I(x, p) \).

(d) The convexity of \( I(\cdot, p) \) and \( I(x, \cdot) \) follows from the convexity of \( u \log u \) and \(-\log u \) for \( u > 0 \).

The next lemma is concerned with \( I(A, p) \), the infimum of \( I(x, p) \) for \( x \in A \). The relevance of \( I(A^{(N)}, p) \) for the approximation of \( P_N(A|p) \) is clear from theorem 2.1. If the set \( A \) is sufficiently regular, the approximation (2.9) is true with \( I(A^{(N)}, p) \) replaced by \( I(A, p) \) (see section 6 and the Appendix).

**Lemma 4.2.** Let \( A \) be a non-empty subset of \( \Omega \).

(a) Let \( p \in \Omega \). Then there is at least one point \( y \) such that

\[
(4.3) \quad y \in \overline{A}, \quad I(y, p) = I(A, p)
\]

If \( p \in \overline{A} \) then \( I(A, p) = 0 \) and (4.3) is satisfied only with \( y = p \). If \( p \notin \overline{A} \) then \( I(A, p) > 0 \) and any \( y \) which satisfies (4.4) is in the boundary of \( A \).

(b) Let \( p \notin \Omega \). Then \( I(A, p) < \infty \) if and only if the intersection \( A \cap \Omega(p) \) is not empty. If this the case, then \( I(A, p) = I(A \cap \Omega(p), p) \) and the statements of part (a) are true with \( A \) replaced by \( A \cap \Omega(p) \).

**Proof.** The lemma follows easily from lemma 4.1. We prove only the last assertion of part (a). Let \( p \in \Omega \), \( p \notin \overline{A} \). Then (since \( I(\cdot, p) \) is continuous) \( I(A, p) > 0 \). Suppose that \( I(y, p) = I(A, p) \) for some \( y \) in the interior of \( A \). Then the point \( z = (1-t)y + t p \) is in \( A \) for some positive \( t < 1 \). Since \( I(\cdot, p) \) is convex,

\[
I(z, p) \leq (1-t)I(y, p) + t I(p, p) = (1-t)I(A, p) < I(A, p)
\]

This contradicts the definition of \( I(A, p) \). Hence any \( y \) which satisfies (4.3) is in the boundary of \( A \).
A maximum likelihood estimate of $p$ under the assumption $p \in \Lambda$ is a point $\hat{p} = \hat{p}(z(N))$ which maximizes $P_N(z(N) | p)$ for $p \in \Lambda$ (or $\Lambda$). From (2.4) we see that $\hat{p}$ minimizes $I(z(N), p)$, so that $I(z(N), \hat{p}) = I(z(N), \Lambda)$. By extension, we may define $\hat{p}(x)$ for any $x \in \Omega$ as a point in $\Lambda$ for which $I(x, \hat{p}(x)) = I(x, \Lambda)$. The next lemma asserts the existence of at least one $\hat{p}(x)$ for each $x$.

**Lemma 4.2.** Let $\Lambda$ be a non-empty subset of $\Omega$. (a) For each $x \in \Omega$ there is at least one point $\hat{p}(x)$ such that

$$I(x, \hat{p}(x)) = I(x, \Lambda).$$

(b) If $x \in \Lambda$ then $I(x, \Lambda) = 0$ and (4.4) is satisfied only for $\hat{p}(x) = x$.

(c) If $x \notin \Lambda$ then $I(x, \Lambda) > 0$ and any $\hat{p}(x)$ which satisfies (4.4) is in the boundary of $\Lambda$.

(d) $I(x, \Lambda)$ is bounded in $\Omega$ if and only if $\Lambda \cap \Omega$ is not empty.

The lemma follows easily from lemma 4.1.

**Remark.** $I(x, \Lambda)$ may be bounded and not continuous in $\Omega$. For example, if $\Lambda$ consists of two points, $\overline{p}^1 = (1, 0, \ldots, 0)$ and $\overline{p}^2 \in \Omega$, then $I(x, \Lambda)$ is discontinuous at $x = \overline{p}^1$.

**Lemma 4.4.** Let $\Lambda$ and $\widehat{\Lambda}$ be non-empty subsets of $\Omega$. Suppose that $I(x, \Lambda)$ is continuous in $\Omega$ and $I(A, \Lambda) > 0$. Then there is at least one point $y$ such that

$$y \in \Lambda, \quad I(y, \Lambda) = I(A, \Lambda),$$

and any point $y$ which satisfies (4.5) is in the boundary of $\Lambda$.

**Proof.** The existence of a point $y$ which satisfies (4.5) follows from the assumed continuity of $I(x, \Lambda)$. Suppose that $I(y, \Lambda) = I(A, \Lambda)$ for some $y$ in the interior of $\Lambda$. By lemma 4.3 there is a $\hat{p} \in \Lambda$ such that $I(y, \hat{p}) = I(y, \Lambda)$. The point $z = (1-t)y + t\hat{p}$ is in $\Lambda$ for some positive $t < 1$. Since $I(\cdot, \hat{p})$ is convex,
\[(4.6)\quad I(z, \hat{p}) \leq (1-t) I(y, \hat{p}) = (1-t) I(A, \Lambda) < I(A, \Lambda),\]
due to \(I(A, \Lambda) > 0\). But since \(\hat{p} \in \Lambda\), since \(I(z, \cdot)\) is continuous, and \(z \in A\), we have \(I(z, \hat{p}) \geq I(z, \Lambda) \geq I(A, \Lambda)\), which contradicts \((4.6)\). This implies the lemma.

We conclude this section with some remarks on the determination of the infimum \(I(A, p)\) and on the set of points in \(\overline{A}\) at which the infimum is attained.

We restrict ourselves to the case \(p \in \Omega_0\). (Lemma 4.2 implies that the general case can be reduced to this case.) The set \(A\) is contained in the set \(\{ x | I(x, p) \geq I(A, p) \}\), whose complement \(C\) is convex. The following lemma gives information about the boundary of \(C\). A hyperplane (briefly: plane) in \(\Omega\) is a non-empty set \(\{ x | \Sigma a_i x_i = c \}\), where \(a_1, \ldots, a_k\) are not all equal. The dimension of a hyperplane is at most \(k-2\); in degenerate cases, such as \(\{ x | x_1 = 1 \}\), the dimension may be less than \(k-2\).

\[4.5\]

**Lemma 4.5.** Let \(p \in \Omega_0\),
\[(4.7)\quad C = \{ x | I(x, p) < c \}\quad , \quad 0 < c < \max_x I(x, p) ,\]
and let \(y\) be a boundary point of \(C\).

(a) if \(y \in \Omega_0\) then
\[(4.8)\quad I(x, p) - c = \sum_{i=1}^{k} (\log (y_i/p_i)) (x_i - y_i) + I(x, y)\]
and the unique tangent plane of \(C\) at \(y\) is \(T = \{ x | \Sigma (\log(y_i/p_i))(x_i - y_i) = 0 \}\).

(b) If \(y \in \Omega_0\) then, for each \(j\) with \(y_j = 0\), \(T_j = \{ x | x_j = 0 \}\) is a tangent plane of \(C\) at \(y\), and \(C\) has no tangent planes at \(y\) other than these \(T_j\) and their intersections.

(c) All boundary points of \(C\) are in \(\Omega_0\) if and only if
\[(4.9)\quad c < -\log (1 - p_{\min}), \quad p_{\min} = \min_i p_i .\]
Proof. (a) The identity (4.8) follows immediately from the fact that
$I(y, p) = c$ and $y$ and $p$ are in $\Omega$. Hence $T$ is a tangent plane of $C$ at $y$. It is unique since the derivatives $\partial I(x, p)/\partial x_i = \log (x_i/p_i) + 1$ are continuous in $\Omega$.

(b) Clearly, if $y_j = 0$, then $T_j = \{x \mid x_j = 0\}$ is a tangent plane of $C$ at $y$. It is sufficient to prove that no hyperplane containing points in $\Omega$ is a tangent plane at $y$. This will follow if we show that every straight line containing $y$ and some point in $\Omega$ intersects the open convex set $C$.

Let $x_0$ be a point in $\Omega$,

$$z(t) = (1-t)y + tx_0,$$

$$F(t) = I(z(t), p).$$

We must show that $F(t) < c$ for some $t \in (0, 1)$.

We have for $t \in (0, 1)$

$$F'(t) = \frac{\partial}{\partial t} I(z(t), p) = \sum_{i=1}^{k} (x_i^0 - y_i) \log (z_i(t)/p_i)$$

$$= \sum_{y_i=0} x_i^0 \log(t x_i^0/p_i) + o(1)$$

$$= \sum_{y_i=0} x_i^0 \log t + o(1) \quad \text{as } t \to 0.$$

Since $F''(t) > 0$, we have by Taylor's formula for $t \in (0, 1)$

$$c = I(y, p) = F(0) = F(t) - t F'(t).$$

It follows that $F(t) < c$ for $t$ positive and sufficiently small, as was to be proved.

(c) Let $x$ be a point with $x_j = 0$. Let $\overline{p}$ be defined by $\overline{p}_j = 0$, $\overline{p}_i = p_i/(1-p_j)$, $i \neq j$. Then

$$I(x, p) = \sum_{i} x_i \log(x_i/\overline{p}_i) - \log (1 - p_j) \geq - \log (1 - p_j),$$

with equality for $x = \overline{p}$. Hence $I(x, p) \leq c$ implies $x \in \Omega$ if, and only if, $c < - \log (1 - p_{\min})$. Part (c) follows.
Lemma 4.6. Let

\[(4.10) \quad A = \{ x \mid f(x) > 0 \} ,\]

where the function \( f(x) \) is continuous in \( \Omega \) and \( \max f(x) > 0 \). Let \( p \) be a point in \( \Omega_o \) such that \( f(p) < 0 \). Suppose further that the derivatives

\[ f_i'(x) = \frac{\partial f(k)}{\partial x_i}, \quad i = 1, \ldots, k, \]

exist and are continuous at all \( x \) in \( \Omega_o \) for which \( f(x) = 0 \).

Let \( y \) be any point in \( \overline{A} \) such that \( I(y, p) = I(A, p) \). Then if \( y \in \Omega_o \), it is necessary that \( f(y) = 0 \) and

\[(4.11) \quad \log \left( \frac{y_i}{p_i} \right) = a f_i'(y) + b \quad i = 1, \ldots, k ,\]

where \( a > 0 \) and \( b \) are constants.

Proof. By lemma 4.2, any point \( y \) in \( \overline{A} \) for which \( I(y, p) = I(A, p) \) is in the boundary of \( A \). Since \( f(x) \) is continuous, this means that \( f(y) = 0 \).

The method of Lagrange multipliers yields the necessary condition \((4.11)\) with some constants \( a \) and \( b \). That \( a \) must be positive follows from

\[ f(x) = f(x) - f(y) = \sum f_i'(y) (x_i - y_i) + o(\|x - y\|) \]

and \((4.8)\) since \( f(x) > 0 \) implies \( I(x, p) > I(y, p) \).

Lemma 4.7. If \( A \) is convex, \( A \cap \Omega_o \) is not empty, and if \( p \in \Omega_o, p \not\in \overline{A} \), then there is exactly one point \( y \in \overline{A} \) such that \( I(y, p) = I(A, p) \).

Proof. The point \( y \) is a common boundary point of the disjoint convex sets \( A \) and \( B' = \{ x \mid I(x, p) < I(A, p) \} \). Since \( A \) and \( B' \) contain points in \( \Omega_o \), it follows from lemma 4.5 (b) that \( y \) is in \( \Omega_o \). Lemma 4.5 (a) implies that the separating hyperplane of the sets \( A \) and \( B' \) is unique, and \( y \) is the unique point in \( \overline{B'} \) which is in that hyperplane.
5. The infimum of $I(x, p)$ subject to the condition $I(x, \Lambda) < c$. The infimum $I(B', p)$, where $B' = \{ x \mid I(x, \Lambda) < c \}$, is needed for the approximation of the power of a likelihood ratio test for testing the hypothesis $p \in \Lambda$. For the case of a simple hypothesis, where $\Lambda$ consists of a single point $p^0$, the problem is solved explicitly (theorem 5.1) and an asymptotic expression for the infimum is obtained (theorem 5.2). The case of an arbitrary $\Lambda$ is then briefly discussed.

**Theorem 5.1.** Let $p^0$ and $p$ be points in $\Omega$, $c$ a finite positive number

\[(5.1) \quad B' = \{ x \mid I(x, p^0) < c \} .\]

(I) We have $I(B', p) < \infty$ if and only if

\[(5.2) \quad \max_i p_i^0 p_i / 0 \quad \text{and} \quad - \log \frac{\sum p_i^0}{p_i} < c .\]

(II) Suppose that condition (5.2) is satisfied. Then there is a unique point $y$ such that

\[(5.3) \quad I(y, p^0) \leq c , \quad I(B', p) = I(y, p) .\]

If $I(\overline{p}, p^0) \leq c$ then $y = \overline{p}$, where

\[(5.4) \quad \overline{p}_i = p_i^0 / \sum_{p_j \neq 0} p_j \quad \text{if} \quad p_i^0 \neq 0 ; \quad \overline{p}_i = 0 \quad \text{if} \quad p_i^0 = 0 .\]

If $c < I(\overline{p}, p^0)$ then

\[(5.5) \quad y_i = (p_i^0)^{1-s} p_i^s / M(s) , \quad i = 1, \ldots, k ,\]

and

\[(5.6) \quad I(B', p) = c - M'(s)/M(s) ,\]

where, for $0 < t < 1$ ,

\[(5.7) \quad M(t) = \sum_{i=1}^k (p_i^0)^{1-t} p_i^t , \quad M'(t) = d M(t)/dt ,\]

and the number $s(0 < s < 1)$ is uniquely determined by

\[(5.8) \quad s \frac{M'(s)}{M(s)} - \log M(s) = c .\]
Proof. First assume that \( p^0 \) and \( p \) are in \( \Omega_0 \). Then \( \overline{p} = p \) and the functions \( I(\cdot, p^0) \) and \( I(\cdot, p) \) are continuous and bounded in \( \Omega \). By lemma 4.2 there is at least one point \( y \) such that (5.3) is satisfied, and if \( I(p, p^0) \leq c \), then necessarily \( y = p \).

Now suppose that \( 0 < c < I(p, p^0) \). Then \( y \) is a common boundary point of the disjoint convex sets \( B' \) and \( C = \{ x \mid I(x, p) < I(B', p) \} \). Since \( B' \) contains points in \( \Omega_0 \), lemma 4.5 (b) implies that \( y \) must be in \( \Omega_0 \). By lemma 4.6 with \( f(x) = -I(x, p^0) + c \) we must have

\[
\log \left( \frac{y_i}{p_i} \right) = -a \log \left( \frac{y_i}{p_i^0} \right) + b, \quad i = 1, \ldots, k,
\]

where \( a > 0 \). This is equivalent to (5.5) with \( s = 1/(1+a) > 0 \). The point \( y \) must satisfy the conditions \( \nabla y_i = 1 \) and \( I(y, p^0) = c \). This implies that \( M(s) \) is given by (5.7) and \( s \) must satisfy (5.8). Thus \( s \) is a positive root of the equation \( F(t) = c \), where

\[
F(t) = t L'(t) - L(t), \quad \quad L(t) = \log M(t).
\]

Now \( F'(t) = t L''(t) > 0 \) for \( t > 0 \). Also, \( F(1) = L'(1) - L(1) = M'(1) = I(p, p^0) \) > \( c \). Hence \( s \) is uniquely determined by (5.8), and \( 0 < s < 1 \).

One easily calculates that \( I(B', p) = I(y, p) \) is equal to the right-hand side of (5.6). This completes the proof for the case where \( p^0 \) and \( p \) are in \( \Omega_0 \).

Now consider the general case. Define \( \overline{p}^0 \) by

\[
\overline{p}_i^0 = \frac{p_i^0}{\sum_{j \neq k} p_j^0} \quad \text{if} \quad p_i \neq 0; \quad \overline{p}_i^0 = 0 \quad \text{if} \quad p_i = 0.
\]

In order that \( I(x, p) \) be finite for some \( x \) such that \( I(x, p^0) < c \), it is necessary that \( x \in \Omega(p^0) \cap \Omega(p) \). If this is the case, then

\[
I(x, p^0) = I(x, \overline{p}^0) + I(\overline{p}^0, p^0) \geq I(\overline{p}^0, p^0) = -\log \sum_{p_i \neq 0} p_i^0.
\]
These facts imply part (I) of the theorem.

If condition (5.2) is satisfied, it follows from the preceding paragraph and the identity

\[ I(x, p) = I(x, \overline{p}) - \log \sum_{p_i^0 \neq 0} p_i \text{ for } x \in \Omega(p^0) \cap \Omega(p) \]

that \( I(B', p) \) is the infimum of \( I(x, \overline{p}) - \log \sum_{p_i^0 \neq 0} p_i \) subject to the conditions \( x \in \Omega(p^0) \cap \Omega(p) \) and \( I(x, \overline{p}) < c + \log \sum_{p_i^0 \neq 0} p_i = \overline{c} \), say. The solution of this problem follows immediately from the first part of the proof, with \( \Omega \), \( p^0 \), \( p \), \( c \) replaced by \( \Omega(p^0) \cap \Omega(p) \), \( \overline{p} \), \( \overline{p} \), \( \overline{c} \). It can be verified that the result is equivalent to that stated in the theorem.

We now derive an asymptotic expression for the infimum \( I(B', p) \) of theorem 5.1 as \( c \to 0 \). We confine ourselves to the case \( p \in \Omega \). In this case, by theorem 5.1, \( I(B', p) \) is finite for small values of \( c \) only if \( p^0 \in \Omega \). To emphasize the dependence on \( c \) we write \( B'(c) \) for \( B' \).

**Theorem 5.2.** Let \( p^0 \in \Omega \), \( p \in \Omega \), \( B'(c) = \{ x \mid I(x, p^0) < c \} \). Then as \( c \to 0 \),

\[ I(B'(c), p) = I(p^0, p) - (2m_2)^{1/2} \frac{3^{1/2}}{s} + (1 + \frac{m_3}{3m_2}) c + o(3^{1/2}) \]

where

\[ m_j = k \sum_{i=1} p_i^0 (\log \frac{p_i^0}{p_i})^j \]

**Proof.** By theorem 5.1,

\[ I(B'(c), p) = c - L'(s_c) \]

where \( L(t) = \log M(t) \), \( M(t) = \sum_{p_i^0} p_i^0 \), \( s_c > 0 \) is determined by

\[ F(s_c) = c \quad F(t) = t L'(t) - L(t) \]

All derivatives of \( L(t) \) and \( F(t) \) exist for all real \( t \), and we have

\[ F'(t) = t L''(t), \quad F''(t) = L''(t) + t L'''(t), \]

\[ F'''(t) = 2L''''(t) + t L^{(4)}(t) \]
Since $F(0) = 0$ and $F(t)$ is strictly increasing for $t > 0$, we have $s_c \to 0$ as $c \to 0$.

As $t \to 0$,

$$F(t) = \frac{1}{2} L''(0) t^2 + \frac{1}{3} L'''(0) t^3 + o(t^4).$$

It is easy to calculate that

$$L(0) = 0, \quad L'(0) = -I(p^0, p), \quad L''(0) = m_2, \quad L'''(0) = -m_3,$$

where $m_j$ is defined by (5.10). Hence as $c \to 0$,

$$c = \frac{1}{2} m_2 s_c^2 - \frac{1}{3} m_3 s_c^3 + o(s_c^4).$$

This implies

$$s_c = \left(\frac{2}{m_2}\right)^{\frac{1}{2}} c^\frac{1}{2} + \frac{2}{3} \left(\frac{m_2}{m_3}\right) c + o(c^{3/2}).$$

Now

$$L'(s_c) = L'(0) + L''(0) s_c + \frac{1}{2} L'''(0) s_c^2 + o(s_c^3)$$

$$= -I(p^0, p) + m_2 s_c - \frac{1}{2} m_3 s_c^2 + o(s_c^3).$$

With (5.16) this yields

$$L'(s_c) = -I(p^0, p) + \left(\frac{2m_2}{m_3}\right)^{\frac{1}{2}} c^{\frac{1}{2}} - \frac{1}{3} \left(\frac{m_3}{m_2}\right) c + o(c^{3/2}).$$

The expansion (5.9) follows from (5.11) and (5.17).

Now let $A$ be a non-empty subset of $\Omega$ and

$$B' = \{ x \mid I(x, A) < c \}.$$ 

We briefly discuss two methods for evaluating $I(B', p)$.

Suppose for simplicity that $p \in \Omega_o$ and assume

$$0 < c < I(p, A).$$

By lemma 4.2 there is at least one point $y$ such that

$$y \in \overline{B'}, \quad I(y, p) = I(B', p),$$

and $y$ must be in the boundary of $B'$. If $I(\cdot, A)$ is continuous, this means that $I(y, A) = c$. Note that in general the set $B'$ is not convex and there may be
more than one minimizing point \( y \).

By lemma 4.3, for each \( y \) there is at least one point \( \hat{p}(y) \) such that

\[
(5.21) \quad \hat{p}(y) \in \Lambda, \quad I(y, \hat{p}(y)) = I(y, \Lambda).
\]

Let

\[
(5.22) \quad B'_y = \{ x \mid I(x, \hat{p}(y)) < I(y, \hat{p}(y)) \}.\]

The following lemma shows that \( I(B'_y, p) = I(B'_y, p) \). For given points \( y \) and \( \hat{p}(y) \) the infimum \( I(B'_y, p) \) is obtained from theorem 5.1. Thus the knowledge of some pair \( (y, \hat{p}(y)) \) enables us to evaluate \( I(B'_y, p) \).

**Lemma 5.1.** Let \( B' \) be defined by (5.18). Suppose that \( p \in \Omega_o \) and condition (5.19) is fulfilled. Let \( y \) and \( \hat{p}(y) \) be points which satisfy (5.20) and (5.21). Then the set \( B'_y \) defined by (5.22) is a subset of \( B' \) and

\[
(5.23) \quad I(B'_y, p) = I(B'_y, p).
\]

**Proof.** Since \( \hat{p}(y) \in \bar{\Lambda} \) and, by lemma 4.3, \( I(x, \Lambda) = I(x, \bar{\Lambda}) \), we have \( I(x, \Lambda) \leq I(x, \hat{p}(y)) \) for all \( x \). Since \( y \in \bar{B'} \), \( I(y, \hat{p}(y)) = I(y, \Lambda) \leq c \). It follows that \( B'_y \subset B' \).

This implies \( I(B'_y, p) \leq I(B'_y, p) \). On the other hand, since

\[
I(y, \hat{p}(y)) \leq c < \infty \quad \text{and} \quad I(\cdot, \hat{p}(y)) \quad \text{is continuous in} \quad \Omega (\hat{p}(y)), \quad \text{we have}
\]

\( y \in \bar{B'} \). Hence \( I(B'_y, p) = I(y, p) \geq I(\bar{B'}_y, p) = I(B'_y, p) \), and (5.23) follows.

The following alternative procedure for evaluating \( I(B'_y, p) \) requires no knowledge of the minimizing points \( y \) and \( \hat{p}(y) \) but involves another minimizing problem. We have

\[
(5.24) \quad B' = \bigcup_{p \in \Lambda} B'(p^0), \quad B'(p^0) = \{ x \mid I(x, p^0) < c \}.
\]

This implies, at least for \( p \in \Omega_o \),

\[
(5.25) \quad I(B'_y, p) = \inf \{ I(B'_y(p^0), p) \mid p^0 \in A \}.
\]

For each \( p^0 \), \( I(B'_y(p^0), p) \) can be obtained from theorem 5.1. Thus the problem of evaluating \( I(B'_y, p) \) is reduced to that of minimizing \( I(B'_y(p^0), p) \) for \( p^0 \in A \).
6. The set of alternatives at which a likelihood ratio test is better than a given test. In this section we shall consider tests which satisfy certain regularity conditions and shall investigate the set of alternatives at which a likelihood ratio test of approximately the same size has a smaller error probability than the given test when $N$ is sufficiently large.

Definition 6.1. A sequence $\{A_N\}$ of subsets of $\Omega$ is said to be regular relative to a point $p$ in $\Omega$ if

$$I(A_N, p) = I(A_N, p) + O(N^{-1} \log N).$$

A sequence $\{A_N\}$ is said to be regular relative to a subset $\Lambda$ of $\Omega$ if

$$I(A_N, \Lambda) = I(A_N, \Lambda) + O(N^{-1} \log N).$$

A subset $A$ of $\Omega$ is said to be regular relative to $p$ (or $\Lambda$) if (6.1) (or (6.2)) holds with $A_N = A$.

Sufficient conditions for a sequence of sets to be regular relative to $p$ or $\Lambda$ are derived in the Appendix.

From theorem 2.1 and definition 6.1 we immediately obtain

Theorem 6.1. If the sequence $\{A_N\}$ is regular relative to $p$ then

$$P_N(A_N | p) = \exp \left\{ -N I(A_N, p) + O(\log N) \right\}.$$

If $\{A_N\}$ is regular relative to $\Lambda$ then

$$\sup_{p \in \Lambda} P_N(A_N | p) = \exp \left\{ -N I(A_N, \Lambda) + O(\log N) \right\}.$$

We now state another version of theorem 3.1 which compares a test $A_N$ for testing the hypothesis $p \in \Lambda$ with a likelihood ratio test. Let

$$B(c) = \{ x | I(x, \Lambda) \geq c \}.$$

Theorem 6.2. Let $\{A_N\}$ be a sequence of sets regular relative to $\Lambda$ and let

$$c_N = I(A_N, \Lambda).$$
There exist positive numbers \( a_N = O(N^{-1} \log N) \) such that

\[
\sup_{p \in \Lambda} P_N \left( B(c_N + a_N) \mid p \right) \leq \sup_{p \in \Lambda} P_N \left( A_N \mid p \right),
\]

and for any \( p \in \Lambda \) such that the sequence \( \{ A_N \} \) is regular relative to \( p \),

\[
P_N \left( B'(c_N + a_N) \mid p \right) \leq \exp \left( -N d_N(p) + N e_N(p) + O(\log N) \right) P_N \left( A_N \mid p \right),
\]

provided that the two probabilities in (6.8) are different from 0, where

\[
d_N(p) = I(B'(c_N), p) - I(A_N, p) \geq 0, \tag{6.9}
\]

\[
e_N(p) = I(B'(c_N), p) - I(B'(c_N + a_N), p) \geq 0. \tag{6.10}
\]

The proof is clear if we note that relations (6.3) and (6.4) with \( \approx \) replaced by \( \leq \) are true for arbitrary sets \( A_N \). Hence to obtain inequalities (6.7) and (6.8) it is not necessary to assume that the sets \( B(c_N + a_N) \) and their complements are regular.

The assumption that the probabilities in (6.8) are positive implies that \( I(A_N, p) \) and \( I(B'(c_N + a_N), p) \) are finite, so that the differences \( d_N(p) \) and \( e_N(p) \) are defined.

By (6.8), if (i) \( N d_N(p)/\log N \to \infty \) and (ii) \( e_N(p)/d_N(p) \to 0 \), then the ratio \( P_N \left( B'(c_N + a_N) \mid p \right)/P_N \left( A_N \mid p \right) \) tends to 0 faster than any power of \( N \). If \( A_N = A \) is independent of \( N \), so are \( c_N = c \) and \( d_N(p) = d(p) \), and conditions (i) and (ii) reduce to \( d(p) > 0 \) and \( e_N(p) \to 0 \). The latter is true if \( I(B'(c), p) \) is a continuous function of \( c \), and then we need only determine the set of points \( p \) for which \( d(p) \) is positive. In the general case the set where \( d_N(p) > 0 \) is also of primary importance, as will be seen in the sequel. The following theorem gives a characterization of this set. To simplify the notation we omit the subscripts \( N \).
Theorem 6.3. Let $A$ and $A'$ be non-empty subsets of $\Omega$ such that $0 < I(A, A') < \infty$. Let

\[(6.11)\quad B = \{ x \mid I(x, A) \geq I(A, A') \}, \]

and for any $p$ such that $I(A', p) < \infty$ let

\[(6.12)\quad d(p) = I(B', p) - I(A', p).\]

(I) Always $d(p) \geq 0$; $d(p) = 0$ if and only if

\[(6.13)\quad \{ x \mid I(x, p) < I(B', p) \} \subset A.\]

(II) If $d(p) = 0$ and $0 < I(B', p) < \infty$ then $I(B', p) = I(y, p)$ for some common boundary point of $y$ of $A$ and $B$.

(III) Suppose that $d(p) = 0$ and $0 < I(B', p) < \infty$. Let $y$ be a common boundary point of $A$ and $B$ such that $I(y, p) = I(B', p)$, and let $\hat{p}(y)$ be a point in $A$ such that $I(y, \hat{p}(y)) = I(y, A)$. If $p$ and $y$ are in $\Omega$ then $\hat{p}(y) \in \Omega$ and $p$ is on the curve

\[(6.14)\quad p = p(t), \quad -\infty < t < 0,\]

where

\[(6.15)\quad p_i(t) = \hat{P}_i(y)^t y_i^{1-t} \sum_{j=1}^{k} \hat{P}_j(y)^t y_j^{1-t}, \quad i = 1, \ldots, k.\]

Proof. (I) Since $A \subset B$, $d(p) \geq 0$. If $d(p) = 0$ then $x \in A'$ implies $I(x, p) \geq I(B', p)$, which is equivalent to (6.13). If (6.13) is satisfied then $x \in A'$ implies $I(x, p) \geq I(B', p)$, hence $I(A', p) \geq I(B', p)$ and therefore $d(p) = 0$.

(II) Suppose that $d(p) = 0$ and $0 < I(B', p) < \infty$. First assume $p \in \Omega$. By lemma 4.2, $I(B', p) = I(y, p)$, where $y$ is in the boundary of $B'$. Since $B' \subset A'$, $y$ is in $A'$, and $I(y, p) = I(A', p)$. Again by lemma 4.2, $y$ is in the boundary of $A'$. Thus $y$ is a common boundary point of $A$ and $B$.

If $p \notin \Omega$, the proof is analogous, with reference to lemma 4.2 (b).
(III) Under the assumptions of part (III) the conditions of lemma 5.1 with
\( c = I(A, \Lambda) \) are satisfied. Hence the set
\[ B'_y = \{ x \mid I(x, \hat{\gamma}(y)) < I(y, \hat{\gamma}(y)) \} \]
is a subset of \( B' \) and \( I(y, p) = I(B', p) = I(B'_y, p) \). (This is true without the assumption \( y \in \Omega_0 \).) Since \( y \in B^c \), we have \( I(y, \hat{\gamma}(y)) = I(y, \Lambda) \leq I(A, \Lambda) < \infty \). Hence \( y \in \Omega (\hat{\gamma}(y)) \). In particular, if \( y \in \Omega_0 \) then \( p(y) \in \Omega_0 \).

It follows from theorem 5.1 with \( p^0 = \hat{\gamma}(y) \) (or, more directly, by an argument used in the proof of that theorem) that
\[ \log(y_i/p_i) = -a \log (y_i/p_i(y)) + b, \quad i = 1, \ldots, k, \]
where \( a > 0 \). This is equivalent to (6.14) and (6.15) with \( t = -a < 0 \). The proof is complete.

Remarks on theorem 6.3. We have excluded the case \( I(A', p) = \infty \), which implies \( I(B', p) = \infty \) and \( P_N(A', p) = P_N(B', p) = 0 \).

If \( I(B', p) = 0 \), that is, \( p \in B^c \), then clearly \( d(p) = 0 \). (In this case the set on the left of (6.13) is empty.) At such alternatives \( p \) the error probabilities \( P_N(A' \mid p) \) and \( P_N(B' \mid p) \) can not be very small.

The alternatives \( p \) of interest to us are those for which \( I(B', p) > 0 \). The conditions \( I(A', p) < \infty \) and \( d(p) = 0 \) imply \( I(B', p) < \infty \). If \( I(\cdot, \Lambda) \) is continuous then \( I(B', p) > 0 \) if and only if \( I(p, \Lambda) > I(A, \Lambda) \). (Note that if \( A = A_N \) depends on \( N \) in such a way that \( I(A_N, \Lambda) \to 0 \) then \( I(p, \Lambda) > I(A_N, \Lambda) \) for each \( p \notin A \) for \( N \) large enough.)

Theorem 6.3 shows that the set of points \( p \) for which \( I(B', p) > 0 \) and \( d(p) = 0 \) essentially depends on the set of common boundary points of \( A \) and \( B \). Suppose, in particular, that the test with critical region \( A^{(N)} \) differs sufficiently from a likelihood ratio test in the sense that the sets \( A \) and \( B \) have only
finitely many common boundary points \( y \). Under some additional conditions theorem 6.3 implies that the set of points \( p \) with \( d(p) = 0 \) is small in a specified sense; this is made precise in the corollary stated below.

To simplify the statement of the theorem we have assumed in part (III) that \( y \) as well as \( p \) are in \( \Omega_o \). For the general case theorem 5.1 implies a similar result except that, for given points \( y \) and \( \hat{p}(y) \), the set where \( d(p) = 0 \) may be of more than one dimension. (Compare example 8.2 in section 8.)

Under the assumptions of part (III) the condition that \( p \) is on the curve (6.14) is necessary but not in general sufficient for \( d(p) = 0 \). It is sufficient if, for instance, the complement \( A' \) of \( A \) is convex.

We state the following implication of theorem 6.3.

**Corollary 6.3.** Let \( 0 < I(A, \Lambda) < \infty \) and let \( B \) and \( d(p) \) be defined by (6.11) and (6.12). Suppose that the number of common boundary points \( y \) of \( A \) and \( B \) is finite; that all these points \( y \) are in \( \Omega_o \); and that for each \( y \) there are only finitely many points \( \hat{p}(y) \in \Lambda \) such that \( I(y, \hat{p}(y)) = I(y, \Lambda) \). Then if \( I(B', p) > 0 \) and \( p \in \Omega_o \), we have \( d(p) > 0 \) except perhaps when \( p \) is on one of the finitely many curves (one for each pair \( (y, \hat{p}(y)) \) defined by (6.14) and (6.15).

In the special case of a simple hypothesis, where \( \Lambda \) consists of a single point \( p^0 \), we have \( B = \{ x \mid I(x, p^0) \geq I(A, p^0) \} \). Here \( p(y) = p^0 \) for all \( y \). If \( p^0 \in \Omega_o \) then, by lemma 4.5, the condition \( I(A, p^0) < - \log (1 - p_{\min}^0) \) is sufficient for all common boundary points of \( A \) and \( B \) to be in \( \Omega_o \).

We conclude this section with a lemma concerning the behavior of \( e_N(p) \) as defined in (6.10) for the case where \( \Lambda \) consists of a single point \( p^0 \).
Lemma 6.1. Let $p^o \in \Omega$, $p \in \Omega$,
\[ B^o(c) = \{ x \mid I(x, p^o) < c \} . \]

Then as $\delta \to 0^+$,
\[ (6.16) \quad I(B^o(c), p) - I(B^o(c + \delta), p) = o(\delta \cdot c^{-\frac{1}{2}}) \]
uniformly for $0 < c < I(p, p^o) - \gamma$, $\gamma > 0$.

Proof. Let $J(c) = I(B^o(c), p)$. By theorem 5.1, $J(c) = c - L'(s_c)$ for $0 < c < I(p, p^o)$, where $0 < s_c < 1$, $F(s_c) = c$, $F(t) = tL'(t) - L(t)$, $L(t) = \log M(t)$, and $M(t)$ is defined in (5.7). For the derivative $s_c' = ds_c/dc$ we have $F'(s_c)s_c' = 1$. Since $F'(t) = tL''(t)$, we obtain $L''(s_c)s_c s_c' = 1$. Hence
\[ J'(c) = 1 - L''(s_c) s_c' = 1 - s_c^{-1} , \]
\[ J''(c) = s_c^{-2} s_c' = s_c^{-3} L''(s_c)^{-1} > 0 . \]

Therefore for $\delta > 0$,
\[ 0 > J(c + \delta) - J(c) \geq 8 J'(c) = 8 \frac{1 - s_c}{s_c} . \]

For $c$ bounded away from 0 and $I(p, p^o)$, $s_c$ is bounded away from 0 and 1. As $c \to 0$, $s_c \approx (2/m_2)^{\frac{1}{2}} c^{\frac{1}{2}}$ by (5.16), where $m_2 > 0$. This implies the lemma.
7. Chi-square and likelihood ratio tests of a simple hypothesis.

Let \( p^0 \) be a point in \( \Omega_0 \). Let

\[
Q^2(x, p^0) = \sum_{i=1}^{k} \frac{(x_i - p^0_i)^2}{p_i^0}.
\]

The chi-square test for testing the simple hypothesis \( H: p = p^0 \) rejects \( H \) if \( Q^2(z(N), p^0) \geq \frac{c^2}{N} \), where \( c_N \) is a positive number. We shall compare this test with the likelihood ratio test which rejects \( H \) if \( I(z(N), p^0) \geq c_N \), where \( c_N \) is so chosen that the two tests have approximately the same size.

It is well known that if \( p = p^0 \) then the random variables \( N Q^2(z(N), p^0) \) and \( 2N I(z(N), p^0) \) have the same limiting \( \chi^2 \) distribution with \( k-1 \) degrees of freedom. Hence if \( 2 \frac{c^2}{N} = c_N = c/N \), where \( c \) is a positive constant, the sizes of the two tests converge to the same positive limit. In fact, in this case the critical regions of the two tests differ very little from each other when \( N \) is large. Indeed, we have

\[
I(x, p^0) = \frac{1}{2} Q^2(x, p^0) + O(\sqrt{N}) \tag{7.2}
\]

where \( |x - p^0| \) denotes the Euclidean distance between \( x \) and \( p^0 \). This implies that the set \( \{x \mid Q^2(x, p^0) < 2c/N\} \) both contains and is contained in a set of the form \( \{x \mid I(x, p^0) < c/N + O(N^{-3/2})\} \).

Hence it can be shown that at any point \( p \neq p^0 \) which is in \( \Omega_0 \) the ratio of the error probabilities of the two tests is bounded away from 0 and \( c/N \) as \( N \to \infty \). (If \( p \notin \Omega_0 \), the error probabilities at \( p \) of both tests are zero for \( N \) sufficiently large.)

In this section it will be shown that if \( c_N \) tends to 0 not too rapidly, then at "most" points \( p \) the error probability of the likelihood ratio test is much smaller than that of the chi-square test when \( N \) is large enough.
We first observe that

\[ Q^2(x, p^o) = \frac{1}{k} \sum_{i=1}^{k} \frac{x_i(x_i - p^o)}{p_i^o} \leq \frac{1}{k} \sum_{i=1}^{k} \frac{x_i(1 - p^o)}{p_i^o} \]

\[ \leq \max \left( 1 - \frac{p_i^o}{p_i} \right) = \frac{1 - p_{\text{min}}^o}{p_{\text{min}}} , \]

where \( p_{\text{min}}^o = \min p_i^o \). The upper bound is attained if and only if \( x_j = 1 \) and \( x_i = 0 \), \( i \neq j \), for some \( j \) such that \( p_j^o = p_{\text{min}}^o \).

Hence when we consider the test defined by \( Q^2(x, p^o) \geq \varepsilon^2 \), we may assume that \( \varepsilon^2 \leq \frac{1 - p_{\text{min}}^o}{p_{\text{min}}} \). The case \( \varepsilon^2 = \frac{1 - p_{\text{min}}^o}{p_{\text{min}}} \) is trivial, and we shall restrict ourselves to the case of strict inequality.

Note that \( p_{\text{min}}^o \leq 1/k \), and \( p_{\text{min}}^o < 1/2 \) unless \( k = 2 \) and \( p^o = (\frac{1}{2}, \frac{1}{2}) \).

Let

\[ A(\varepsilon) = \{ x \mid Q^2(x, p^o) \geq \varepsilon^2 \} \].

Theorem 7.1. Suppose that \( p^o \in \Omega_o \) and

\[ 0 < \varepsilon < \left( \frac{1 - p_{\text{min}}^o}{p_{\text{min}}} \right)^{\frac{1}{2}} . \]

If \( r \) is the number of components \( p_j^o \) such \( p_j^o = p_{\text{min}}^o \), there are exactly \( r \) points \( y \) such that

\[ y \in A(\varepsilon), \quad I(A(\varepsilon), p^o) = I(y, p^o) . \]

These points are defined by

\[ y_i^C = b \frac{p_{\text{min}}}{p^o} \text{ if } i = j \text{ and } p_j^o = p_{\text{min}}^o, \quad y_i^C = a \frac{p_i^o}{p^o} \text{ if } i \neq j , \]

\[ a = 1 - \left( \frac{p_{\text{min}}^o}{1 - p_{\text{min}}^o} \right)^{\frac{1}{2}} \varepsilon, \quad b = 1 + \left( \frac{1 - p_{\text{min}}^o}{p_{\text{min}}^o} \right)^{\frac{1}{2}} \varepsilon , \]

and we have

\[ I(A(\varepsilon), p^o) = p_{\text{min}}^o b \log b + (1 - p_{\text{min}}^o) a \log a . \]

Furthermore,
\[(7.10) \quad 2p^0_{\text{min}} (1 - p^0_{\text{min}}) \xi^2 \leq \frac{p^0_{\text{min}}(1-p^0_{\text{min}})}{1-2p^0_{\text{min}}} \log \frac{1-p^0_{\text{min}}}{p^0_{\text{min}}} \xi^2 \leq I(A(t), p^0) \leq \varepsilon^2, \]

where the second expression is to be replaced by \(\frac{1}{2} \xi^2\) if \(p^0_{\text{min}} = \frac{1}{2}\). As \(t \to 0\),

\[(7.11) \quad I(A(\varepsilon), p^0) = \frac{1}{2} \xi^2 + \frac{1}{6} \frac{2p^0_{\text{min}} - 1}{(p^0_{\text{min}}(1-p^0_{\text{min}}))} \xi^3 + O(\varepsilon^4). \]

**Proof.** Let \(y = y^\xi\) denote any point which satisfies (7.6). By lemma 4.2 we must have \(Q^2(y, p^0) = \xi^2\). It can be shown that necessarily \(y \in \Omega_\varepsilon\). (For \(\varepsilon\) small enough, \(Q^2(y, p^0) = \xi^2\) implies \(y \in \Omega_\varepsilon\). In general this result can be deduced from lemma 4.5(b). The details are left to the reader.) By lemma 4.6 we must have

\[\log (y_i/p^0_1) = s y_i/p^0_1 + t, \quad i = 1, \ldots, k,\]

where \(s > 0\). Hence \(y_i/p^0_1\) can take at most two different values, say

\[(7.12) \quad y_i = a p^0_1 \quad \text{if} \quad i \in M, \quad y_i = b p^0_1 \quad \text{if} \quad i \notin M,\]

where \(M\) is a non-empty proper subset of \(\{1, \ldots, k\}\). The conditions \(\Sigma y_i = 1\) and \(Q^2(y, p^0) = \xi^2\) imply

\[(7.13) \quad a h + b(1-h) = 1, \quad (a-1)^2 h + (b-1)^2 (1-h) = \xi^2.\]

\[(7.14) \quad h = \Sigma p^0_1. \quad \text{if} \quad i \in M.\]

We may assume \(a < b\). Then

\[(7.15) \quad a = 1 - ((1-h)/h)^\frac{1}{2} \xi, \quad b = 1 + (h/(1-h))^{\frac{1}{2}} \xi.\]

To satisfy \(y_1 > 0\) we must have \(a > 0\), that is, \(\xi^2 < h/(1-h)\). If \(\xi^2\) is close to its upper bound \((1-p^0_{\text{min}})/p^0_{\text{min}}\), this condition is satisfied only when \(h\) takes its largest possible value, \(1-p^0_{\text{min}}\). It will be shown that, for any \(\varepsilon\), \(y\) satisfies (7.6) if and only if \(h = 1 - p^0_{\text{min}}\).

For \(y\) defined by (7.12) we have
I(y, p^o) = h a \log a + (1 - h)b \log b = f(h), say, where a = a(h) and b = b(h) are given by (7.15). By a straightforward calculation we obtain for the derivative of f(h)

\[ f'(h) = b \left( 1 - \frac{a}{b} + \frac{1}{2} \left( 1 + \frac{a}{b} \right) \log \frac{a}{b} \right). \]

The expression on the right is negative. Hence as h ranges over the values (7.14), f(h) attains its minimum at \( h = 1 - p^o \). This implies that condition (7.6) is satisfied if and only if \( y^c \) is one of the points defined by (7.7) and (7.8), and that \( I(A(\xi), p^o) \) is given by (7.9).

The inequality \( I(A(\xi), p^o) \leq \varepsilon^2 \) in (7.10) follows from the general inequality

\[ (7.16) \quad I(x, p) = \sum x_i \log(x_i/p_i) \leq \sum x_i ((x_i/p_i) - 1) = Q^2(x, p). \]

The first two inequalities in (7.10) are contained in theorem 1 of Hoeffding [5]. (Note that the closer lower bound in (7.10) is attained for \( \varepsilon = (1 - 2p_{\min})/(p_{\min}^o (1 - p_{\min}^o))^{1/2} \).

The expansion (7.11) is easily verified. The proof is complete.

The next theorem gives the infimum of \( I(x, p) \) subject to the condition \( x \in A'(\xi) \), that is, \( Q^2(x, p^o) < \varepsilon^2 \).

**Theorem 7.2.** Let \( p^o \in \Omega, p \in \Omega, \xi > 0, \varepsilon^2 < Q^2(p, p^o) \). Then there is unique point \( z^c \) such that

\[ (7.17) \quad I(A'(\xi), p) = I(z^c, p). \]

The point \( z^c \) is determined by the conditions

\[ (7.18) \quad \frac{z^c_i - p^o_i}{p^o_i} = -s_i \log(z^c_i/p^o_i) + t_i, \; i = 1, \ldots, k; \quad s_i > 0, \]

\[ (7.19) \quad \sum_{i=1}^{k} z^c_i = 1, \quad Q^2(z^c, p^o) = \varepsilon^2. \]

As \( \varepsilon \to 0 \),
Proof. By lemma 4.7 there is exactly one point $z^\xi$ which satisfies (7.17). By lemma 4.6, $z^\xi$ must satisfy (7.18) with $s_\xi > 0$. The constants $s_\xi$ and $t_\xi$ are determined by (7.19).

Now let $\xi \to 0$. The condition $Q^2(z^\xi, p^0) = \xi^2$ implies $z^\xi - p^0_i = O(\xi)$, $i = 1, \ldots, k$. Hence

$$\log \left( \frac{z^\xi_i}{p^0_i} \right) = \log \left( \frac{p^0_i}{p^0_i} \right) + (z^\xi_i - p^0_i)/p^0_i + O(\xi^2).$$

With (7.18) this gives

$$(7.23) \quad (1 + s_\xi) \left( \frac{z^\xi_i}{p^0_i} - \frac{p^0_i}{p^0_i} \right) / p^0_i = t_\xi - s_\xi (\log \left( \frac{p^0_i}{p^0_i} \right) + O(\xi^2)).$$

Multiplication of both sides with $p^0_i$ and summation yields

$$(7.24) \quad t_\xi = s_\xi \left( I(p^0, p) + O(\xi^2) \right).$$

From (7.23) and (7.24) we obtain

$$(7.25) \quad \left( z^\xi_i - p^0_i \right) / p^0_i = - \frac{s_\xi}{1 + s_\xi} \left( \log \left( \frac{p^0_i}{p^0_i} \right) - I(p^0, p) + O(\xi^2) \right).$$

If we square both sides of (7.25), multiply with $p^0_i$ and sum with respect to $i$, we find that

$$\xi^2 = \left( \frac{s_\xi}{1 + s_\xi} \right)^2 (m_2(p) + O(\xi^2)).$$

Hence

$$(7.26) \quad \frac{s_\xi}{1 + s_\xi} = m_2(p)^{-\frac{1}{2}} \xi + O(\xi^3).$$

From (7.25) and (7.26) we obtain relation (7.20).
(7.27) \( I(z^{'}, p) = I(p^0, p) + \sum (z^{'i} - p^0i) \log (p^0i/p_i) + \frac{1}{2} \xi^2 + O(\xi^3) \).

With (7.17) and (7.20) this implies (7.21). This completes the proof.

Let \( A(\xi) \) be defined by (7.4) and let

\[
B(\xi) = \{ x \mid I(x, p^0) \geq I(A(\xi), p^0) \},
\]

\[
d(p, \xi) = I(B(\xi), p) - I(A(\xi), p).
\]

**Theorem 7.3.** Let \( p^0 \in \Omega_0 \), \( 0 < \xi < ((1-p^0o/p^0o)^{\frac{1}{2}} \).

(I) If \( p \in \Omega_0 \) and \( \xi^2(p, p^0) > \xi^2 \) then \( d(p, \xi) > 0 \) unless for some \( j \)

with \( p^0_j = p^0 \)

\[
p_j = 1 - a + a p_j; \quad p_i = a p^0_i, \quad i \neq j,
\]

\[
0 < a < 1 - (p^0o/(1 - p^0o))^{\frac{3}{2}} \xi^2.
\]

(II) As \( \xi \to 0 \),

\[
d(p, \xi) = \frac{1}{6} m_2(p)^{\frac{1}{2}} \Delta(p) \xi^2 + O(\xi^3),
\]

where \( m_j(p) \) is defined in (7.22) and

\[
\Delta(p) = \frac{m_2(p)}{m_2(p)^{\frac{3}{2}}} + \frac{1 - 2p^0o}{p^0o(1 - p^0o)} \frac{1/2}{(p^0o(1 - p^0o))^{\frac{3}{2}}}.
\]

(III) We have \( \Delta(p) \geq 0 \) for all \( p \in \Omega_0, \) \( p \neq p^0; \) and \( \Delta(p) > 0 \) unless \( p \) satisfies (7.30) with \( a \neq 1, \) \( 0 < a < (1 - p^0o)^{-1} \) for some \( j \) such

that \( p^0_j = p^0 \).

**Proof.** Part (I) follows from theorems 6.3 and 7.1. The parameter \( a \) replaces the parameter \( t \) in theorem 6.3.

(II) By theorem 7.1, as \( \xi \to 0 \),

\[
I(A(\xi), p^0) = \frac{1}{2} \xi^2 + \frac{1}{6} \frac{2p^0o - 1}{(p^0o(1 - p^0o))^{\frac{3}{2}}} \xi^3 + O(\xi^4).
\]
By theorem 5.2,  
\[(7.35) \ I(B(\xi), p) = I(p^0, p) - (2m_2(p))^{1/2} c^{3/2} + (1 + \frac{m_2(p)}{2m_2(p)})c + O(c^{3/2}),\]
where \( c = I(A(\xi), p^0). \) From (7.34) and (7.35) we obtain after simplification  
\[(7.36) \ I(B(\xi), p) = I(p^0, p) - m_2(p)^{1/2} \xi \]
\[+ \left\{ \frac{1}{2} + \frac{m_2(p)}{6m_2(p)} - \frac{1}{6} m_2(p)^{3/2} \right\} c^2 + O(c^3).\]
By theorem 7.2,  
\[(7.37) \ I(A(\xi), p) = I(p^0, p) - m_2(p)^{1/2} \xi + \frac{1}{2} \xi^2 + O(\xi^3).\]
The expression (7.32) for \( d(p, \xi) \) follows from (7.36) and (7.37).

(III) Let \( u = (u_1, \ldots, u_k), \quad \mu_j(u) = \sum_{i=1}^k \frac{p^0_i u_i}{u_j}. \)

Then \( \mu_1(u) = 0, \quad \mu_2(u) = 1, \quad \mu_j(u) = m_2(p)/m_2(p)^{3/2}. \) Part (III) of theorem 7.3 is an immediate consequence of the following lemma. (Note that \( \Delta(p) \geq 0 \) is implied by \( d(p, \xi) \geq 0. \) The lemma gives the conditions for equality.)

**Lemma 7.1.** Let \( \mu_j(u) \) be defined by (7.38), where \( p^0 \in \Omega_0 \) and \( u_1, \ldots, u_k \) are any real numbers such that \( \mu_1(u) = 0 \) and \( \mu_2(u) = 1. \) Then  
\[(7.39) \ \mu_3(u) \geq \frac{(2p^0_{\min} - 1)/(p^0_{\min} - p^0_{\min})}{\xi^2}.\]
The sign of equality holds if and only if for some \( j \) such that \( p^0_j = p^0_{\min} \)  
\[(7.40) \ u_j = -((1 - p^0_{\min})/p^0_{\min})^{1/2}; \quad u_i = (p^0_{\min}/(1 - p^0_{\min}))^{1/2}, \quad i \neq j.\]

**Proof.** Since \( p^0 \in \Omega_0, \) the set of points \( u \) defined by \( \mu_1(u) = 0, \mu_2(u) = 1 \) is bounded and closed. Hence \( \mu_3(u) \) has a finite minimum in this set. An application of the method of Lagrange multipliers shows that for \( u \) to be a minimizing point it is necessary that \( u_1 \) take only two values, say \( u_1 = a \) if \( i \in M, \quad u_1 = b \) if \( i \notin M, \quad a > b. \) The conditions \( \mu_1(u) = 0, \mu_2(u) = 1 \) imply \( \mu_2(u) = (1 - 2h)/(h(1 - h))^{1/2}, \) where \( h = \sum_{i \in M} \frac{p^0_i}{p^0_{\min}}. \) The minimum with
respect to $h$ of this ratio is attained at $h = 1 - P_{\text{min}}^0$, and the lemma follows.

The following lemma establishes the regularity of the sequences of sets $\{A(\xi_N)\}$ and $\{A'\xi_N\}$ under general conditions.

**Lemma 7.2.** Let $p^0 \in \Omega_0$, $A(\xi) = \{x \mid Q^2(x, p^0) \geq \xi^2\}$. For any $\xi_N$, the sequence $\{A(\xi_N)\}$ is regular at $p^0$. If $p \in \Omega_0$, $\xi^2_N < Q^2(p, p^0)$, and

$$N^2 \xi^2_N \rightarrow \infty \quad \text{as} \quad N \rightarrow \infty,$$

the sequence $\{A'(\xi_N)\}$ is regular at $p$.

**Proof.** Since $A'(\xi)$ is convex, the sequence $\{A(\xi_N)\}$ is regular at $p^0$ by theorem A.1 of the Appendix.

By theorem 7.2, if $\xi^2 < Q^2(p, p^0)$ then $I(A'(\xi), p) = I(z^*, p)$, where $z^*$ is defined in the theorem. Theorem A.2 of the Appendix implies that $\{A'(\xi_N)\}$ is regular at $p$ if

$$N \max_{i,j} \left| \frac{(\xi_{Ni}^0)}{p_{i}^0} - \frac{(\xi_{Nj}^0)}{p_{j}^0} \right| \rightarrow \infty \quad \text{as} \quad N \rightarrow \infty.$$

This condition is satisfied if $\xi_N$ is bounded away from 0 or, by (7.20), if $N \xi_N \rightarrow \infty$. The lemma follows.

Remark. The case where condition (7.41) is not satisfied is of no statistical interest since if $N^2 \xi^2_N$ is bounded, the size of the test tends to 1. If $\xi^2_N \leq a^2/N^2$ and $a$ is sufficiently small, the set $A'(\xi_N)^{(N)}$ is empty for infinitely many $N$ and hence the sequence $\{A'(\xi_N)\}$ is not regular at any $p$ in $\Omega_0$.

We now can state the following result about the relative performance of a chi-square test and a likelihood ratio test of a simple hypothesis.

**Theorem 7.4.** Let $p^0 \in \Omega_0$ and $0 < \xi_N < \left(\frac{(1-P_{\text{min}}^0)}{P_{\text{min}}^0}\right)^\frac{1}{2}$.

(I) For the error probabilities of the chi-square test which rejects the hypothesis $p = p^0$ if $z(N) \in A(\xi_N) = \{x \mid Q^2(x, p^0) \geq \xi^2_N\}$ we have...
(7.43) \( P_N(A(\xi_N) \mid p^0) = \exp \{ -N I(A(\xi_N), p^0) + O(\log N) \} \),

where \( I(A(\xi), p^0) \) is given explicitly in theorem 7.1; and if

(7.44) \( p \in \Omega_o, \xi^2_N < Q^2(p, p^0), \ N^2 \xi^2_N \to \infty \) as \( N \to \infty \),

then

(7.45) \( P_N(A'(\xi_N) \mid p) = \exp \{ -N I(A'(\xi_N), p) + O(\log N) \} \),

where \( I(A'(\xi), p) \) is given in theorem 7.2.

(II) There exist positive constants \( \alpha_N = O(N^{-1} \log N) \) such that for the likelihood ratio test which rejects \( p = p^0 \) if

\[ z^{(N)} \in R_N = \{ x \mid I(x, p^0) \geq I(A(\xi_N), p^0) + \alpha_N \} \]

we have

(7.46) \( P_N(R_N \mid p^0) \leq P_N(A(\xi_N) \mid p^0) \),

and if conditions (7.44) are satisfied,

(7.47) \( P_N(R^1_N \mid p) \leq \exp \{ -N d(p, \xi_N) + O(\log N/\xi^2 (A(\xi_N), p^0)) \} \ P_N(A'(\xi_N) \mid p) \),

where \( d(p, \xi_N) \) is defined in (7.29) and has the properties stated in theorem 7.3.

In particular, if

(7.48) \( \xi_N^3 \to 0, \ N \xi^3_N/\log N \to \infty \) as \( N \to \infty \),

then at each point \( p \in \Omega_o, p \neq p^0 \) which does not lie on one of the line segments

(7.49) \( p_j = 1 - a + a p_j^0; \ p_i = a p_i^0, i \neq j; \ 0 < a < 1; \ p_j^0 = p_{\text{min}}^0 \),

the likelihood ratio test \( R_N \) is more powerful than the chi-square test \( A(\xi_N) \)

when \( N \) is sufficiently large.

Proof. Part (I) follows from theorem 6.1 and lemma 7.2.

Part (II) follows from theorem 6.2, lemma 6.1, and theorems 7.1 and 7.3.

Remark. The line segments (7.49) connect the point \( p^0 \) with some of the vertices of the simplex \( \Omega \).

It would be of interest to investigate whether for moderately large values of \( N \) the likelihood ratio test is more powerful than the chi-square test except in a
small neighborhood of these line segments when the size of the two tests is not unreasonably small.
8. Chi-square and likelihood ratio tests of a composite hypothesis.

There is reason to believe that in the case of a composite hypothesis the relation between a chi-square test and a likelihood ratio test in general is analogous to that in the case of a simple hypothesis (see section 7), with a notable exception mentioned below. For a chi-square test of a composite hypothesis the determination of the common boundary points of the sets A and B (in the notation of theorem 6.3) is somewhat cumbersome. We therefore present no general results. We first shall show by an example that for one common version of the chi-square test it may happen that the size of the test is never smaller than some power of N; if this is the case, our theory is not applicable. We then give a simple example where the situation is analogous to the case of a simple hypothesis.

There are several versions of the chi-square test for testing a composite hypothesis, \( p \in \Lambda \). One is the minimum chi-square test which is based on the statistic \( Q^2(z^{(N)}, \Lambda) \), where

\[
Q^2(x, \Lambda) = \inf \{ Q^2(x, p) \mid p \in \Lambda \}.
\]

Here \( Q^2(x, p) \) is defined by (7.1), with the convention that \( (x_i - p_i)^2/p_i = 0 \) if \( x_i = p_i = 0 \). The calculation of \( Q^2(x, \Lambda) \) is cumbersome for some of the common hypotheses. When a maximum likelihood estimator \( p(x) \) of \( p \) under the assumption \( p \in \Lambda \) is available (as defined in lemma 4.3) one often resorts to the test based on

\[
\hat{Q}^2(x) = Q^2(x, \hat{p}(x)).
\]

If the size of the test is held fixed as \( N \) increases and the set \( \Lambda \) is sufficiently regular, the tests based on \( Q^2(x, \Lambda) \) or \( \hat{Q}^2(x) \) differ little from a likelihood ratio test based on \( I(x, \Lambda) \), just as in the case of a simple hypothesis. However, if we require that the size of the test tend to 0 more rapidly than a certain power of \( N \), it turns out that this requirement can not in general be satisfied with a \( \hat{Q}^2 \) test.
Let $\hat{\chi}^2_N$ denote the maximum of $\hat{\chi}^2(z^{(N)})$ for $z^{(N)} \in \Omega^{(N)}$. The $\hat{\chi}^2$ test of smallest positive size for testing the hypothesis $H: p \in \Lambda$ rejects $H$ if and only if $\hat{\chi}^2(z^{(N)}) = \hat{\chi}^2_N$. Suppose that this critical region contains a point $z^{(N)}$ which is close to $\Lambda$ in the sense that $I(z^{(N)}, \Lambda) = I(z^{(N)}, p(z^{(N)}))$ is of order $N^{-1} \log N$. Then, by (3.3), the smallest positive size of a $\hat{\chi}^2$ test is not smaller than some power of $N$. The following example serves to illustrate this phenomenon.

Example 8.1: Hypothesis of independence in a contingency table. Let the $k = rs$ components of $x \in \Omega$ be denoted by $x_{ij}$, $i = 1, \ldots, r; j = 1, \ldots, s$, where $r \geq 2, s \geq 2$. Define $x_1^{(1)} = \sum_j x_{ij}$, $x_j^{(2)} = \sum_i x_{ij}$. Let

$$\Lambda = \{ p | p_{ij} = p_i^{(1)} p_j^{(2)}, i = 1, \ldots, r; j = 1, \ldots, s \} .$$

Then $p_{ij}(x) = x_i^{(1)} x_j^{(2)}$ and

$$\hat{\chi}^2(x) = \sum_{i=1}^r \sum_{j=1}^s \frac{x_{ij}^2}{x_i^{(1)} x_j^{(2)}} - 1,$$

where, by definition, the terms with $x_i^{(1)} x_j^{(2)} = 0$ are zero. For simplicity let $r = s$. Then, due to $x_{ij}^2 \leq x_{ij} x_{ij}^2$, $\hat{\chi}^2(x) \leq r-1$, with equality holding if and only if each row and each column of the matrix $x_{ij}$ contains exactly one non-zero element. Let $z^{(N)}$ denote the point defined by $z_{ii}^{(N)} = 1 - (r-1)/N$; $z_{ii}^{(N)} = 1/N$, $i = 2, \ldots, r$; $z_{ij}^{(N)} = 0$, $i \neq j$. Then $\hat{\chi}^2(z^{(N)}) = r - 1$ and

$$P_N(z^{(N)} | p(z^{(N)})) = \frac{N!}{(N-r+1)!} \left(1 - \frac{r-1}{N}\right)^{2(N-r+1)} N^{-2(r-1)} \sim e^{-2r+2} N^{-r+1} .$$

Thus the smallest positive size of a $\hat{\chi}^2$ test is proportional to $N^{-r+1}$.

In contrast, the size of the likelihood ratio test which rejects $H$ if $I(z^{(N)}, \Lambda) \geq c$ does not exceed $\exp \{-Nc + 0 (\log N)\}$.
In the following example the situation is similar to the case of a simple hypothesis.

Example 8.2. Let $x = (x_1, \ldots, x_k)$, $k \geq 3$, 

$$A = \{ p \mid p_1 = p_2 \}.$$ 

We have $p_1(x) = (x_1 + x_2)/2$, $i = 1, 2$; $p_2(x) = x_1$, $i > 2$. Hence

$$\hat{Q}^2(x) = 4(x_1 + x_2) \left( \frac{x_1}{x_1 + x_2} - \frac{1}{2} \right)^2,$$

max $\hat{Q}^2(x) = 1$. Let $A = \{ x \mid Q^2(x) \geq \xi^2 \}$, $0 < \xi < 1$. It can be shown that the sets $A$ and $B = \{ x \mid I(x, A) \geq I(A, A) \}$ have exactly two common boundary points, $y_1 = ((1-\xi)/2, (1+\xi)/2, 0, \ldots, 0)$, $y_2 = ((1-\xi)/2, (1+\xi)/2, 0, \ldots, 0)$. Since these points are not in $\Omega_0$, part (III) of theorem 6.3 is not directly applicable. It is not difficult to show that if $I(p, A) > I(A, A)$ then $d(p) > 0$ unless $p_1 = 0$ or $p_2 = 0$. Thus if $k \geq 4$, the set of points $p$ such that $I(B', p) > 0$ and $d(p) = 0$ is of more than one dimension. Since, however, the present hypothesis set $A$ is such that we have effectively $k = 3$ components, the result may be said to be analogous to that in the case of a simple hypothesis.
9. Some competitors of the chi-square test. There are a number of test statistics for testing a simple hypothesis which have the same asymptotic distribution as the chi-square statistic when the hypothesis is true. As an example consider the test which rejects the hypothesis \( p = p^0 \) if \( D^2_1(z(N), p^0) \) exceeds a constant, where

\[
D^2_1(x, p) = \sum_{i=1}^{k} (x_i - p_i)^2
\]

(see Matusita [6]).

For \( p^0 \in \Omega_0 \) we have

\[
D^2_1(x, p^0) = \frac{1}{4} \Phi^2(x, p^0) + 0(|x - p^0|^3).
\]

Thus if the size of the test is bounded away from zero, the test behaves asymptotically as the chi-square test and differs little from the likelihood ratio test.

Let \( A = \{ x | D^2_1(x, p^0) \geq \epsilon^2 \} \), \( 0 < \epsilon^2 < 2 \). It is easily seen that there are only finitely many points \( y \) in \( A \) for which \( \Phi(y, p^0) = \Phi(A, p^0) \). Just as in the case of the chi-square test they are such that the ratio \( y_i/p_i^0 \) takes only two different values. If the size of the test tends to 0 at an appropriate rate, the test compares with the likelihood ratio test in a similar way as the chi-square test.

For testing a composite hypothesis we may use the test based on \( D^2_1(x, \hat{p}(x)) \). It can be shown that in the case of example 8.1 the size of this test may decrease at an exponential rate, in contrast to the analogous chi-square test.

Another interesting class of tests is defined in terms of the distances

\[
D(x, p) = \max_{M \in \mathcal{M}} \sum_{i \in M} (x_i - p_i),
\]

where \( \mathcal{M} \) is a family of subsets of \( \{1, \ldots, k\} \). If \( \mathcal{M} \) contains all these subsets then \( D(x, p) = \frac{1}{2} \sum |x_i - p_i| \). If \( \mathcal{M} \) consists of the sets \( \{i\} \) and \( \{i, \ldots, k\} \) for \( i = 1, \ldots, k \), then \( D(z(N), p^0) \) may be identified with the Kolmogorov statistic (discrete case).
Let \( p^o \in \Omega_0, \ 0 < \xi < \max_x D(x, p^o) \). The set \( A = \{ x \mid D(x, p^o) \geq \xi \} \) is the union of the half-spaces \( A_M = \{ x \mid \sum_{i \in M} (x_i - p^o_i) \geq \xi \} , M \in \mathcal{M} \). Hence we obtain

\[
I(A, p^o) = \min_{M \in \mathcal{M}} I(A_M, p^o) = \min_{M \in \mathcal{M}} J(h_M)
\]

where \( h_M = \sum_{i \in M} p^o_i \) and

\[
J(h) = (h + \xi) \log \frac{h + \xi}{h} + (1 - h - \xi) \log \frac{1 - h - \xi}{1 - h}
\]

Again the minimizing points \( y \) are such that the ratio \( y_i / p^o_i \) takes only two values. The function \( J(h) \) has a unique minimum at a point \( h_0 \) which is close to \( 1/2 \) if \( \xi \) is small. This implies that if \( D(x, p^o) = \frac{1}{2} \sum |x_i - p^o_i| \) and \( p^o_i = 1/k, i = 1, \ldots, k \), then, for \( \xi \) small, there are close to \( (k/2) \) minimizing points. For the chi-square test this number if only \( k \).
10. Bayes tests and likelihood ratio tests. In this section it will be shown that certain Bayes tests differ little from the corresponding likelihood ratio test if \( N \) is large, not only when the size \( \alpha_N \) of the test is bounded away from 0 (in which case a chi-square test has a similar property) but also when \( \alpha_N \) tends to zero.

Let \( G \) be a distribution function on the simplex \( \Omega \) and let

\[
P_N(z^{(N)} | G) = \int \frac{P_N(z^{(N)} | p)}{dG(p)}.
\]

The Bayes test for testing the hypothesis \( H: p = p^0 \) against the alternative that \( p \) is distributed according to \( G \) rejects \( H \) if the ratio \( P_N(z^{(N)} | G) / P_N(z^{(N)} | p^0) \) exceeds a constant. This ratio is \( \exp \left\{ -N I(z^{(N)}, p^0) \right\} \).

Let \( U \) denote the uniform distribution on \( \Omega \), so that the vector \((p_1, \ldots, p_{k-1})\) has a constant probability density. We have

\[
P_N(z^{(N)} | U) = \binom{N + k - 1}{k - 1}^{-1}
\]

for all \( z^{(N)} \). Hence

\[
I(z^{(N)}, p^0) - \log \frac{P_N(z^{(N)} | U)}{P_N(z^{(N)} | p^0)} = \log \binom{N + k - 1}{k - 1} + \log P_N(z^{(N)} | z^{(N)}) \cdot
\]

Here the left side is the difference between the test statistics for the likelihood ratio test and the Bayes test. An application of Stirling's formula to the last term in (10.3) (see (2,11)) shows that if the components of \( z^{(N)} \) are bounded away from 0, the right side of (10.3) is of the form \( c_N + O(1) \), where \( c_N \) does not depend on \( z^{(N)} \). This implies that the critical regions of the two tests (when they are of approximately the same size) and their error probabilities at the points in \( \Omega \) differ little from each other.

The uniform distribution \( U \) has been chosen for simplicity. We obtain a similar result if \( U \) is replaced by a distribution \( G \) such that, for example, the probability density of \((p_1, \ldots, p_{k-1})\) is positive and bounded.
Now consider a composite hypothesis, \( H: \theta \in \Lambda \). Let \( \Theta_0 \) be a distribution on \( \Omega \) such that the set \( \Lambda \) has probability one. We may expect that for suitable choices of \( \Theta_0 \) and \( \Theta \) the Bayes test based on the ratio
\[
P_N(z(N) | \Theta_0) / P_N(z(N) | \Theta_0)
\]
will differ little from the likelihood ratio test based on \( I(z(N), \Lambda) \). This is here illustrated by two examples.

**Example 10.1:** Binomial hypothesis. Let \( k = m+1 \) and denote the points of \( \Omega \) by \( x = (x_0, x_1, \ldots, x_m) \). Let
\[
A = \{ \theta(\theta) \mid 0 \leq \theta \leq 1 \}, \quad \pi_i(\theta) = \left( \begin{array}{c} m \\ i \end{array} \right) \theta^i (1 - \theta)^{m-i}, \quad i = 0, 1, \ldots, m.
\]
Then \( I(x, \Lambda) = I(x, \hat{\theta}(x)) \), where \( \hat{\theta}(x) = \pi(\hat{\theta}(x)) \), \( \hat{\theta}(x) = \sum_i x_i / m \).

Let \( U_0 \) denote the distribution on \( \Lambda \) induced by the uniform distribution of \( \theta \) on \( (0,1) \). Then
\[
P_N(z(N) | U_0) = \frac{N!}{\prod_i n_i!} \frac{1}{(mN+1)} \left( \frac{1}{s} \right)^m \prod_{i=0}^{m-1} \left( \begin{array}{c} mN+1 \\ s \end{array} \right),
\]
where \( s = \sum_i n_i = mN \theta(z(N)) \). Let, as before, \( U \) be the uniform distribution on \( \Omega \). After simplification we obtain
\[
N I(z(N), \Lambda) - \log \frac{P_N(z(N) | U)}{P_N(z(N) | U_0)} = \log \left( \frac{m^m}{m} \right) - \log (mN+1) + \log P_N(z(N) | \lambda(N)) - \log P_N^* \,
\]
where
\[
P_N^* = \left( \begin{array}{c} mN \\ s \end{array} \right) \left( \frac{s}{mN} \right)^s \left( 1 - \frac{s}{mN} \right)^{mN-s}.
\]
Relation (10.5) is analogous to (10.3) and implies a similar conclusion.

**Example 10.2:** Hypothesis of independence in a contingency table. Let \( \Lambda \) be defined as in example 8.1. Let \( U_0 \) be the distribution on \( \Lambda \) such that the random vectors \( (\pi_1^{(1)}, \ldots, \pi_r^{(1)}) \) and \( (\pi_1^{(2)}, \ldots, \pi_s^{(2)}) \) are independent and each
is uniformly distributed on the respective probability simplex. Let \( z^{(N)} = n_{ij}/N \), 
\[ n_i^{(1)} = \sum_j n_{ij}, \quad n_j^{(2)} = \sum_i n_{ij}. \]
We obtain
\[
N I(z^{(N)}, \Lambda) - \log \frac{P_N(z^{(N)} \mid U)}{P_N(z^{(N)} \mid U_o)}
\]
(10.7)
\[
= \log \frac{(N+r-s-1)}{(N+r-s)(N+s-1)} + \log \frac{P_N(z^{(N)} \mid z^{(N)})}{P_N(z^{(N)} \mid z^{(N)})},
\]
where
\[
P_N^{(1)} = \frac{N!}{N^{\sum_i n_i^{(1)}}} \prod_{i=1}^{n_i^{(1)}} n_i^{(1)}!
\]
\[P_N^{(2)} \]
is defined in an analogous way in terms of the \( n_j^{(2)} \).

The result is quite similar to that of example 10.1.

The hypothesis sets \( \Lambda \) of examples 10.1 and 10.2 are special cases of a class of subsets of \( \Omega \) for which relations analogous to (10.5) and (10.7) hold true.
Appendix: Regular sequences of sets. In this appendix sufficient conditions are derived for a sequence of subsets of $\Omega$ to be regular relative to a point in $\Omega$. We recall that, by definition 6.1, the sequence $\{A_N\}$ is regular relative to $p$ if

$$I(A_N(p), p) = I(A_N, p) + O(N^{-1} \log N).$$

(Sanov [2] considered the weaker regularity condition where the remainder term in (A.1) is replaced by $o(1)$.)

Since $A_N(\mathcal{N}) \subseteq A_N$, we have $I(A_N(\mathcal{N}), p) \geq I(A_N, p)$. Hence for those $N$ for which $I(A_N, p) = \infty$ condition A.1 is satisfied. Thus $\{A_N\}$ is regular relative to $p$ if condition (A.1), with $=$ replaced by $\leq$, is fulfilled for those sets $A_N$ for which $I(A_N, p) < \infty$.

**Lemma A.1.** The sequence $\{A_N\}$ is regular relative to $p$ if there exist constants $N_0$ and $c$ such that for each $N > N_0$ with $I(A_N, p) < \infty$ there is a point $y \in \Omega$ for which

$$I(y, p) \leq I(A_N, p)$$

and a point $z \in A_N(\mathcal{N})$ for which

$$|z_i - y_i| < c N^{-1} \text{ if } p_i > 0, \quad z_i = 0 \text{ if } p_i = 0.$$

**Proof.** It is sufficient to show that

$$I(z, p) - I(y, p) \leq O(N^{-1} \log N).$$

The assumptions imply that $y$ and $z$ are in $\Omega(p)$ for $N \geq N_1 \geq N_0$. Hence for $N \geq N_1$

$$I(z, p) - I(y, p) = \sum_{p_i \neq 0} d_i, \quad d_i = z_i \log \left(\frac{z_i}{p_i}\right) - y_i \log \left(\frac{y_i}{p_i}\right).$$

If $z_i = 0$ then $y_i < c N^{-1}$ and $d_i = -y_i \log \left(\frac{y_i}{p_i}\right) = O(N^{-1} \log N)$. 
If \( z_i \neq 0 \) then \( z_i \geq N^{-1} \) and

\[
d_i = (z_i - y_i) \log \left( \frac{z_i}{p_i} \right) + y_i \log \left( \frac{z_i}{y_i} \right)
\]

\[
\leq (z_i - y_i) \log \left( \frac{z_i}{p_i} \right) + y_i \left( (z_i/y_i) - 1 \right)
\]

\[
\leq |z_i - y_i| \log N^{-1} + O(|z_i - y_i|) = O(N^{-1} \log N).
\]

Hence \( d_i \leq O(N^{-1} \log N) \) for all \( i \) with \( p_i \neq 0 \), and (A.4) follows.

For any real numbers \( a_1, \ldots, a_k, c \) the subset of \( \Omega \) defined by \( \sum a_i x_i > c \) or \( \sum a_i x_i \geq c \) will be called a **half-space**. (It is convenient here not to exclude the case where all \( a_i \) are equal. Thus the proof of the next lemma for the case \( p \notin \Omega \) is strictly analogous to the proof for \( p \in \Omega \).

**Lemma A.2.** If \( p \) is any point in \( \Omega \) and \( A \) is any half-space such that \( I(A, p) < \infty \), then there is a point \( y \in \Omega \) for which \( I(y, p) = I(A, p) \), and for each \( N \geq k(k-1) \) there is a point \( z \in \mathbb{A}^N \) such that \( |z_i - y_i| \leq (k-1)N^{-1} \) if \( p_i > 0 \) and \( z_i = 0 \) if \( p_i = 0 \).

**Proof.** First assume that \( p \in \Omega \) and

\[
A = \{ x \mid \sum a_i x_i > c \}.
\]

Since \( I(A, p) < \infty \), \( A \) is not empty, so that \( \max a_i > c \). By Lemma 4.2 there is a point \( y \) such that \( I(y, p) = I(A, p) \) and \( \sum a_i y_i \geq c \). It is easy to show that \( y \in \Omega \).

We have \( y_i \geq k^{-1} \) for some \( i \). For definiteness assume that \( y_k \geq k^{-1} \). Define \( z = (z_1, \ldots, z_k) \) as follows. For \( i = 1, \ldots, k-1 \) let \( Nz_i \) be an integer such that

\[
(A.5) \quad z_i \geq 0, \quad z_i \neq y_i, \quad |z_i - y_i| \leq N^{-1}, \quad (a_i - a_k)(z_i - y_i) \geq 0.
\]
These conditions can be satisfied since \( y \in \Omega_0 \). Let \( z_k = 1 - z_1 - \ldots - z_{k-1} \). Then
\[
z_k \geq 1 - \sum_{i=1}^{k-1} (y_i + N^{-1}) = y_k - (k-1) N^{-1} \geq k^{-1} - (k-1) N^{-1} .
\]

Hence if \( N \geq k(k-1) \) then \( z_k \geq 0 \) and \( z \in \Omega(N) \). Moreover,
\[
|z_i - y_i| \leq (k-1) N^{-1} \text{ for all } i .
\]

Now
\[
(A.6) \quad \sum_{i=1}^{k} a_i z_i - c \geq \sum_{i=1}^{k} a_i z_i - \sum_{i=1}^{k} a_i y_i = \sum_{i=1}^{k-1} (a_i - a_k)(z_i - y_i) .
\]

If the \( a_i \) are not all equal, the last sum is strictly positive by (A.5). Otherwise the inequality in (A.6) is strict. Hence \( z \in A(N) \) for \( N \geq k(k-1) \). The lemma is proved for the present case.

If \( p \in \Omega_0 \) and \( A = \{ x \mid \sum a_i x_i > c \} \), the conclusion of the lemma follows from the first part of the proof provided that the set \( \{ x \mid \sum a_i x_i > c \} \) is not empty. If it is empty then, since \( A \) must be non-empty, we have \( \max a_i = c \) and \( A \) is the set of all points \( x \) such that \( x_i = 0 \) if \( a_i < c \).

We have \( I(A, p) = I(y, p) \) where \( y_i = p_i / \sum_{a_j = c} p_j \) if \( a_i = c \), \( y_i = 0 \) otherwise. It is trivial to show that the conclusion of the lemma is true in this case.

If \( p \notin \Omega_0 \), the assumption \( I(A, p) < \infty \) implies \( I(A, p) = I(A \cap \Omega(p), p) \) and the proof is similar to the case \( p \in \Omega_0 \).

Lemmas A.1 and A.2 imply that any sequence of half-spaces is regular relative to any point in \( \Omega \). More generally we have

**Theorem A.1.** Any sequence of subsets of \( \Omega \) whose complements are convex is regular relative to any point in \( \Omega \).

**Proof.** Let \( A \) be a set whose complement is convex and \( p \) a point such that \( I(A, p) < \infty \). We restrict ourselves to the case \( p \in \Omega_0 \) since the case \( p \notin \Omega_0 \) is treated in an analogous way, as in the proof of lemma A.2.
Let $y$ be a point in $A$ such that $I(y, p) = I(A, p)$. Since $A'$ is convex, $A$ is the union of half-spaces bounded by the supporting hyperplanes of $A'$.

If $y \in \Omega_o$ then $y$ is in the closure of one of the non-empty half-spaces $H$ whose union is $A$, and $I(A, p) = I(H, p)$. By lemma A.2 for each $N \geq k(k-1)$ there is a point $z$ in $H(N)$, hence in $A(N)$, with the properties there stated.

If $y \notin \Omega_o$ then the set $A \cap \Omega(y)$ is not empty; this follows from the convexity of $A'$ and lemma 4.5. We have $I(A, p) = I(A \cap \Omega(y), p)$. The set $A' \cap \Omega(y)$ is a convex subset of $\Omega(y)$. For $x \in \Omega(y)$ we have $I(x, p) = I(x, p) - \log \Sigma' p_j$, where $\Sigma'$ denotes the sum over the $j$ with $y_j \neq 0$ and $p_j = p_j / \Sigma' p_j$ if $y_j \neq 0$, $p_j = 0$ otherwise. The argument of the preceding paragraph, with $\Omega$ replaced by the subspace $\Omega(y)$, leads to the conclusion of that paragraph.

Thus the conclusion of lemma A.2 is true for any subset $A$ of $\Omega$ whose complement is convex. With lemma A.1 this implies the theorem.

Define the subset $\Omega_{\xi}$ of $\Omega$ by

$$\Omega_{\xi} = \{ x \mid x_i > \xi, \ i = 1, \ldots, k \} .$$

**Theorem A.2.** Let

$$A_N = \{ x \mid f(x) > c_N \} ,$$

where the $c_N$ are real numbers and $f(x)$ is a function defined on $\Omega$ whose derivatives $f'_i(x) = \partial f(x) / \partial x_i$ and $f''_{ij}(x) = \partial^2 f(x) / \partial x_i \partial x_j$ exist and are continuous in $\Omega_o$. Let $p \in \Omega_o$. Suppose that there exist positive numbers $N_o$ and $\xi$ and for each $N > N_o$ there is a point $y^{(N)}$ such that

$$y^{(N)} \in \bigcap_{\xi} f(y^{(N)}) = c_N, \quad I(y^{(N)}, p) = I(A_N, p)$$

and
(A. 10) \[ \lim_{N \to \infty} N \max_{i,j} \left| f_i^1(y(N)) - f_j^1(y(N)) \right| = +\infty. \]

Then the sequence \( \{ A_N \} \) is regular relative to \( p \).

**Proof.** The assumptions imply that for \( N > N_0 \)

\[ f(x) - c_N = f(x) - f(y(N)) \]
\[ = \sum_{i=1}^{k} f_i^1(y(N)) (x_i - y_i(N)) + O(\|x-y(N)\|^2) \]
\[ = \sum_{i=1}^{k-1} a_i(N) (x_i - y_i(N)) + O(\|x-y(N)\|^2) \]

uniformly for \( x \in \Omega_{r/2} \), where \( a_i(N) = f_i^1(y(N)) - f_i^1(y(N)) \).

For \( i = 1, \ldots, k-1 \) let \( m_i(N) \) denote the largest integer \( \leq N y_i(N) \) and let
\( m_k(N) = N - m_1(N) - \cdots - m_{k-1}(N) \). Define the point \( z(N) \) by
\[ N z_i(N) = m_i(N) + 2 \text{ if } a_i(N) \geq 0, \quad N z_i(N) = m_i(N) - 1 \text{ if } a_i(N) < 0, \]
for \( i \leq k-1 \) and \( z_k(N) = 1 - z_1(N) - \cdots - z_{k-1}(N) \). Then

\[ (A.11) \quad |z_i(N) - y_i(N)| < \frac{2k}{N}, \quad i = 1, \ldots, k. \]

Since \( y(N) \in \Omega_r \), \( z(N) \) is in \( \Omega_{r/2} \) for \( N > N_1 \geq N_0 \). Moreover,
\[ a_i(N) (y_i(N) - y_i(N)) \geq N^{-1} |a_i(N)|, \quad i = 1, \ldots, k-1. \]

Hence for \( N > N_1 \)
\[ f(z(N)) - c_N \geq N^{-1} \sum_{i=1}^{k-1} |a_i(N)| + O(N^{-2}) \]
\[ \geq \frac{1}{2} N^{-1} \max_{i,j} \left| f_i^1(y(N)) - f_j^1(y(N)) \right| + O(N^{-2}). \]

Condition (A.10) implies that for \( N \) large enough we have \( f(z(N)) > c_N \), that is \( z(N) \in A_N \).

Thus the conditions of lemma A.1 are satisfied. The proof is complete.
REFERENCES


369. Ray, S. N. Some sequential Bayes procedures for comparing two binomial parameters when observations are taken in pairs. 1963.
385. Khatri, C. G. Distribution of the largest or smallest characteristic root. 1964.
388. Khatri, C. G. Monotonicity of the power functions of some tests for a particular kind of multicollinearity and unbiasedness and restricted monotonicity of the power functions of the step down procedures for MANOVA and independence. 1964.
393. Burton, R. C. An application of convex sets to the construction of error correcting codes and factorial designs. 1964.
394. Roy, S. N. A report on some results in simultaneous or joint linear (point) estimation. 1964.
399. Khatri, C. G. On a MANOVA model applied to problems in growth curve. 1964.