THE LIKELIHOOD RATIO TEST FOR THE MULTINOMIAL DISTRIBUTION

J. OOSTERHOFF
UNIVERSITY OF NIJMEGEN
and
W. R. VAN ZWET
UNIVERSITY OF LEIDEN

1. Introduction and summary

Let $X^{(N)} = (X_1^{(N)}, \cdots, X_k^{(N)})$ be a random vector having a multinomial distribution with parameters N and $p = (p_1, \cdots, p_k)$,

$$(1.1) \qquad P(X^{(N)} = x|p) = \frac{N!}{x_1! \cdots x_k!} p_1^{x_1} \cdots p_k^{x_k},$$

where $x = (x_1, \cdots, x_k)$ is a vector with nonnegative integer components with sum N, and p is any point in the simplex

$$(1.2) \qquad \Omega = \left\{(y_1, \cdots, y_k) \middle| \sum_{i=1}^{k} y_i = 1, y_i \geqq 0 \text{ for } i = 1, \cdots, k\right\}.$$

By $Z^{(N)} = (Z_1^{(N)}, \cdots, Z_k^{(N)})$ we denote the random vector with components

$$(1.3) \qquad Z_i^{(N)} = \frac{X_i^{(N)}}{N}, \qquad i = 1, \cdots, k.$$

For $N = 1, 2, \cdots$, consider tests based on $Z^{(N)}$ for the hypothesis $H: p \in \Lambda_0$ against the alternative $K: p \in \Lambda_1$, where Λ_0 and Λ_1 are disjoint subsets of Ω and $\Lambda = \Lambda_0 \cup \Lambda_1$ may be a proper subset of Ω. It is assumed that the sizes α_N of the tests depend on N in such a way that $\alpha_N \to 0$ for $N \to \infty$. The likelihood ratio test based on $Z^{(N)}$ for H against K rejects H for large values of the statistic

$$(1.4) \qquad \inf_{p \in \Lambda_0} \sup_{\pi \in \Lambda} \sum_{i=1}^{k} Z_i^{(N)} \log \frac{\pi_i}{p_i},$$

possibly with randomization on the set where the statistic assumes its critical value.

In [2] W. Hoeffding considered a special case of this situation where $\Lambda = \Omega$, in which case the likelihood ratio statistic (1.4) reduces to

Report SW 3/70 Mathematisch Centrum, Amsterdam.

$$\inf_{p \in \Lambda_0} \sum_{i=1}^{k} Z_i^{(N)} \log \frac{Z_i^{(N)}}{p_i}. \tag{1.5}$$

The paper [2] is devoted to making precise the following proposition in this case: *"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 p in the set of alternatives when N is large enough, provided that $\alpha_N \to 0$ at a suitable rate."* By "considerably more powerful" is meant that the ratio of the error probabilities of the second kind at p of the two tests tends to zero more rapidly than any power of N. The condition that "$\alpha_N \to 0$ at a suitable rate" will typically imply that α_N tends to zero more rapidly than any power of N, that is, that $-\log \alpha_N / \log N \to \infty$.

If the likelihood ratio test is much better than a given test for most alternatives, it is natural to ask how much worse it can be for the remaining alternatives or sequences of alternatives. Let β_N denote the power function of the size α_N likelihood ratio test based on $Z^{(N)}$ for H against K and let β_N^+ be the size α_N envelope power for testing H, that is, $\beta_N^+(p)$ is the power at p of the size α_N most powerful test based on $Z^{(N)}$ for H against the simple alternative p. The shortcoming of the size α_N likelihood ratio test for a given N is defined by

$$R_N(p) = \beta_N^+(p) - \beta_N(p), \qquad p \in \Lambda_1. \tag{1.6}$$

The main purpose of this paper is to show that for a simple hypothesis H and under a condition concerning the speed of convergence of α_N to zero, the shortcoming of the likelihood ratio test converges to zero uniformly on the set of alternatives. We note that for testing the simple hypothesis $H: p = p^0$, $p^0 \in \Lambda$ against $K: p \in \Lambda_1 = \Lambda - \{p^0\}$ the likelihood ratio statistic (1.4) reduces to

$$\sup_{\pi \in \Lambda} \sum_{i=1}^{k} Z_i^{(N)} \log \frac{\pi_i}{p_i^0}. \tag{1.7}$$

THEOREM 1.1. *Let Λ be an arbitrary subset of Ω, p^0 an arbitrary point of Λ and let R_N denote the shortcoming of the size α_N likelihood ratio test based on $Z^{(N)}$ for $H: p = p^0$ against $K: p \in \Lambda_1 = \Lambda - \{p^0\}$. If*

$$\lim_{N\to\infty} \alpha_N = 0, \qquad -\log \alpha_N = o(N) \qquad \text{for} \quad N \to \infty, \tag{1.8}$$

then

$$\lim_{N\to\infty} \sup_{p \in \Lambda_1} R_N(p) = 0. \tag{1.9}$$

Although Hoeffding's result and Theorem 1.1 are complementary in the sense mentioned above, we wish to point out that they are of an entirely different nature. Hoeffding's theorem concerns fixed alternatives and the performance of the likelihood ratio test is compared to that of a fixed sequence of tests by considering the ratio of error probabilities of the second kind. The alternatives at which the likelihood ratio test is considerably more powerful in Hoeffding's sense are necessarily alternatives where the power of the likelihood ratio test tends to one very rapidly. Since also the convergence of α_N to zero is assumed to be fast, the probabilities to be considered under the hypothesis as well as under the alternative are all probabilities of large deviations. The tools used to estimate these probabilities are Theorems 2.1 and A.1 in [2] which are reproduced here as Lemma 2.6.

In Theorem 1.1 on the other hand the performance of the likelihood ratio test is compared at each alternative to that of the most powerful test for that alternative. The comparison is in terms of power difference and the result is uniform on the set of alternatives. Alternatives or sequences of alternatives for which the power of the likelihood ratio test tends to one play a role only in so far as uniformity is concerned and the theorem is basically concerned with sequences of alternatives for which the power of the likelihood ratio test remains bounded away from one. Under alternatives we only have to compute probabilities of small deviations which is done by applying the central limit theorem. As α_N is allowed to tend to zero either slowly or fast, we are dealing with intermediate as well as large deviations under the hypothesis. In the former case where $-\log \alpha_N = o(N^{1/6})$, Theorem 1.1 was first proved by using classical limit theorems by J. Oosterhoff in [3] under the additional assumptions that $\Lambda = \Omega$ and that p^0 is an interior point of Ω. We shall use this result (Lemma 3.1) as a starting point for our investigation in the case where α_N tends to zero slowly. In the case where α_N tends to zero fast the resulting probabilities of large deviations are dealt with in the same manner as is done in [2].

The condition $-\log \alpha_N = o(N)$ in Theorem 1.1 is unduly restrictive and occurs there only for the sake of simplicity. In fact we shall show that it may be replaced by the assumption that there exists $\varepsilon > 0$ such that for all sufficiently large N

$$\alpha_N \geqq (1 - p_m^0)^N e^{N\varepsilon}. \tag{1.10}$$

where p_m^0 is the smallest positive coordinate of p^0. Moreover, further refinements of this condition are possible.

The reason that we need an assumption of this type at all, is to avoid complications arising from the fact that under sequences of alternatives converging sufficiently fast to certain boundary points of Ω, the distribution of the likelihood ratio statistic degenerates too rapidly. The nature of these complications is most easily made clear for alternatives located at the extreme points of Ω (that is, the points with a coordinate equal to one).

EXAMPLE 1.1. Take for p^0 the point with coordinates $p_i^0 = k^{-1}, i = 1, \cdots, k$, and suppose that Λ contains all extreme points of Ω. Choose $\alpha_N = k^{-N}$. The statistic (1.7) assumes its maximum value if $Z_i^{(N)} = 1$ for some i. Since $P(Z_i^{(N)} = 1 | p^0) = k^{-N}$ for each i, the size α_N likelihood ratio test rejects $H: p = p^0$ with probability k^{-1} if $Z_i^{(N)} = 1$ for some i and hence its power at each of the extreme points of Ω is equal to k^{-1}. For each i, the size α_N most powerful test for $H: p = p^0$ against the simple alternative $p_i = 1$ rejects H if $Z_i^{(N)} = 1$ and has power one at $p_i = 1$. The shortcoming of the likelihood ratio test at each of the extreme points of Ω is therefore equal to $1 - k^{-1}$ for every N.

It is of course easy to modify this example in such a way that no randomization occurs.

Whereas Hoeffding's result is restricted to the case where $\Lambda = \Omega$ but allows a composite hypothesis H, Theorem 1.1 places no restriction on Λ but deals only with a simple hypothesis H. In Section 4 we shall show by means of a counterexample that even for the case where $\Lambda = \Omega$ Theorem 1.1 does not hold in general for a composite hypothesis H.

Section 2 of this paper contains some preliminary results on the multinomial distribution. In Section 3 we prove Theorem 1.1 and show that the condition $-\log \alpha_N = o(N)$ may be replaced by (1.10). Section 4 is devoted to the case where the hypothesis H is composite.

2. Preliminary results

For any set $A \subset \Omega$ we shall denote by A^N the set of all $y \in A$ for which Ny has integer coordinates.

LEMMA 2.1. *For any $A \subset \Omega$ for which A^N is nonempty, the function $f(p = P(Z^{(N)} \in A|p)$ assumes its maximum value only at points p in the convex hull of A^N.*

PROOF. Let π be a point in the complement of the convex hull of A^N. Since A^N contains only finitely many points its convex hull is closed and hence there exists a hyperplane separating π and A^N, that is, there exists a vector $a = (a_1, \cdots, a_k)$ such that $\Sigma\, a_i(z_i - \pi_i) > 0$ for all $z \in A^N$. Because $\Sigma\, z_i = \Sigma\, \pi_i = 1$, we may choose a in such a way that $\Sigma\, a_i \pi_i = 0$ and $\Sigma\, a_i z_i > 0$ for all $z \in A^N$. As $\Sigma\, a_i \pi_i = 0$ and $a_i = \pi_i = 0$ whenever $\pi_i = 0$, the points with coordinates $\pi_i + \varepsilon a_i \pi_i$ are points of Ω for all sufficiently small $\varepsilon > 0$. Hence

$$\begin{aligned}(2.1)\qquad \sum_{i=1}^{k} a_i \pi_i \frac{\partial}{\partial p_i} f(p)\big|_{p=\pi} &= \sum_{i=1}^{k} a_i \sum_{z \in A^N} P(Z^{(N)} = z|\pi)\, N z_i \\ &= N \sum_{z \in A^N} P(Z^{(N)} = z|\pi) \sum_{i=1}^{k} a_i z_i\end{aligned}$$

is a directional derivative of f at π in a direction in Ω multiplied by a nonnegative constant. Note, however, that $a_i \pi_i$ may be equal to zero for all i if $\pi_i = 0$ for some i.

If $f(\pi) > 0$, then (2.1) is positive because $\Sigma\, a_i z_i > 0$ for all $z \in A^N$ and consequently f does not have a maximum at π. If $f(\pi) = 0$ the same conclusion holds since A^N is nonempty. *Q.E.D.*

For $z, p \in \Omega$ we define

$$(2.2)\qquad I(z, p) = \sum_{i=1}^{k} z_i \log \frac{z_i}{p_i},$$

where $z_i \log (z_i/p_i) = 0$ by definition if $z_i = 0$. It is well known that for fixed p this function is convex in z, positive unless $z = p$ and finite if $p_i \neq 0$ for all i. In Lemma 2.2 we show that under p the random variable $I(Z^{(N)}, p)$ is of order at most N^{-1} in probability uniformly in p.

LEMMA 2.2. *For every $\varepsilon > 0$ there exists $A > 0$ such that for all N*

$$\sup_{p\in\Omega} P\left(I(Z^{(N)}, p) \geqq \frac{A}{N}\Big| p\right) \leqq \varepsilon. \tag{2.3}$$

PROOF. For $0 \leqq z_i \leqq 1$, $0 < p_i \leqq 1$,

$$z_i \log \frac{z_i}{p_i} = z_i \log\left(1 + \frac{z_i - p_i}{p_i}\right) \leqq z_i \frac{z_i - p_i}{p_i} = (z_i - p_i) + \frac{(z_i - p_i)^2}{p_i}. \tag{2.4}$$

Since uncer p, $Z_i^{(N)} = Z_i^{(N)} \log (Z_i^{(N)}/p_i) = 0$ a.s. if $p_i = 0$, we have under p

$$0 \leqq I(Z^{(N)}, p) \leqq \sum_{p_i \neq 0} \frac{(Z_i^{(N)} - p_i)^2}{p_i} \tag{2.5}$$

with probability one. It follows that

$$E(I(Z^{(N)}, p)|p) \leqq \sum_{p_i \neq 0} \frac{p_i(1 - p_i)}{Np_i} \leqq \frac{k-1}{N}. \tag{2.6}$$

Application of Markov's inequality completes the proof.

Let $\mathring{\Omega}$ denote the interior of Ω,

$$\mathring{\Omega} = \{(y_1, \cdots, y_k) | \sum_{i=1}^{k} y_i = 1, y_i > 0 \text{ for } i = 1, \cdots, k\} \tag{2.7}$$

and define for $p^0 \in \mathring{\Omega}$, $p \in \Omega$,

$$\sigma^2(p, p^0) = \sum_{i=1}^{k} p_i \left(\log \frac{p_i}{p_i^0}\right)^2 - \left(\sum_{i=1}^{k} p_i \log \frac{p_i}{p_i^0}\right)^2. \tag{2.8}$$

We shall have to consider the asymptotic distribution of

$$T_p^{(N)} = \sum_{i=1}^{k} Z_i^{(N)} \log \frac{p_i}{p_i^0} \tag{2.9}$$

under p for fixed $p^0 \in \mathring{\Omega}$ and varying $p \in \Omega$. The distribution of $T_p^{(N)}$ under p is degenerate if and only if the positive coordinates of p are proportional to the corresponding coordinates of p^0 (as before we take $0 \log 0 = 0$ by definition). For $p \neq p^0$ and $p \geqq \varepsilon > 0$ (that is, $p_i \geqq \varepsilon$ for all $i = 1, \cdots, k$) the following lemma provides a uniform normal approximation. By Φ we denote the standard normal distribution function.

LEMMA 2.3. *For any fixed $p^0 \in \mathring{\Omega}$ and $\varepsilon > 0$,*

$$\lim_{N\to\infty} P\left(\frac{T_p^{(N)} - I(p, p^0)}{(p, p^0)} N^{1/2} \leqq a | p\right) = \Phi(a) \tag{2.10}$$

uniformly for all a and all $p \in \Omega$ with $p \neq p^0$ and $p \geqq \varepsilon$.

PROOF. Under p the distribution of $NT_p^{(N)}$ is the same as that of $\Sigma_{j=1}^N Y_j$, where $Y_1, \cdots, Y_N$ are independent and identically distributed random variables with

$$P\left(Y_j = \log \frac{p_i}{p_i^0}\right) = p_i, \qquad i = 1, \cdots, k. \tag{2.11}$$

Hence

$$E(T_p^{(N)}|p) = I(p, p^0), \tag{2.12}$$

$$\sigma^2(T_p^{(N)}|p) = N^{-1}\sigma^2(p, p^0). \tag{2.13}$$

Let $F_{N,p}$ be the distribution function of

$$\frac{T_p^{(N)} - I(p, p^0)}{\sigma(p, p^0)} N^{1/2} \tag{2.14}$$

and for $m = 2, 3$, let $\nu_{m,p}$ denote the mth absolute central moment of Y_j. Since the distribution of Y_j is degenerate only if the positive coordinates of p are proportional to the corresponding coordinates of p^0, $\nu_{m,p}$ is positive and finite if $p \neq p^0$ and $p \in \mathring{\Omega}$. Hence by the Berry-Esseen theorem (see [1]) we have for all a and N and for all $p \in \mathring{\Omega}$, $p \neq p^0$,

$$|F_{N,p}(a) - \Phi(a)| \leqq c\nu_{3,p}\nu_{2,p}^{-3/2} N^{-1/2}, \tag{2.15}$$

where c is a constant independent of a, N and p. By (2.11)

$$\nu_{m,p} = \sum_{j=1}^{k} p_j \eta_j^m, \qquad \eta_j = \left| \log \frac{p_j}{p_j^0} - \sum_{i=1}^{k} p_i \log \frac{p_i}{p_i^0} \right|; \tag{2.16}$$

if $p \neq p^0$ and $p \geqq \varepsilon$ then $p_t\eta_t^3 = \max_j p_j \eta_j^3$ is positive and finite and as a result

$$\nu_{3,p}\nu_{2,p}^{-3/2} \leqq kp_t\eta_t^3 (p_t\eta_t^2)^{-3/2} = kp_t^{-1/2} \leqq k\varepsilon^{-1/2}. \tag{2.17}$$

Together with (2.15) this proves the lemma.

LEMMA 2.4. *For every fixed* $p^0 \in \mathring{\Omega}$ *and* $\varepsilon > 0$ *there exist* $0 < M_1 < M_2 < \infty$ *such that*

$$M_1 I(p, p^0) \leqq \sigma^2(p, p^0) \leqq M_2 I(p, p^0) \tag{2.18}$$

for all $p \in \Omega$ *with* $p \geqq \varepsilon$.

PROOF. By expanding the logarithms involved we find that for $p \in \mathring{\Omega}$ with $\max|p_i - p_i^0| < \delta$,

$$\begin{aligned} I(p, p^0) &= \frac{1}{2} \sum_{i=1}^{k} \frac{(p_i - p_i^0)^2}{p_i^0} + O(\delta^3), \\ \sigma^2(p, p^0) &= \sum_{i=1}^{k} \frac{(p_i - p_i^0)^2}{p_i^0} + O(\delta^3). \end{aligned} \tag{2.19}$$

The proof is completed by noting that for p outside a neighborhood of p^0 and $p \geqq \varepsilon$, both $I(p, p^0)$ and $\sigma^2(p, p^0)$ are bounded away from zero and infinity.

For $p^0 \in \Lambda \subset \Omega$ we shall have to consider

$$\sup_{\pi \in \Lambda} \sum_{i=1}^{k} Z_i^{(N)} \log \frac{\pi_i}{p_i^0}, \tag{2.20}$$

where $z_i \log(\pi_i/p_i^o) = 0$ by definition if $z_i = 0$. Note that under $p \in \Lambda$ this random variable is defined (possibly $+\infty$) with probability one.

LEMMA 2.5. *Let Λ be an arbitrary subset of Ω, p^0 an arbitrary point of Λ and define $\Lambda_1 = \Lambda - \{p^0\}$. Furthermore, let c_N and a_N, $N = 1, 2, \cdots$, be sequences of nonnegative real numbers such that*

$$\lim_{N\to\infty} c_N = 0, \qquad \lim_{N\to\infty} Nc_N = \infty, \qquad \lim_{N\to\infty} \frac{Na_N^2}{c_N} = 0. \tag{2.21}$$

Then

$$\sup_{p\in\Lambda_1} P\left(\sup_{\pi\in\Lambda} \sum_{i=1}^{k} Z_i^{(N)} \log \frac{\pi_i}{p_i^0} \leqq c_N + a_N, I(Z^{(N)}, p^0) \geqq c_N - a_N \middle| p\right) \tag{2.22}$$

tends to zero for $N \to \infty$.

PROOF. Under $p \in \Lambda_1$,

$$\sup_{\pi\in\Lambda} \sum_{i=1}^{k} Z_i^{(N)} \log \frac{\pi_i}{p_i^0} \geqq \sum_{i=1}^{k} Z_i^{(N)} \log \frac{p_i}{p_i^0} = I(Z^{(N)}, p^0) - I(Z^{(N)}, p) \tag{2.23}$$

a.s. since under p, $0 \leqq I(Z^{(N)}, p) < \infty$ a.s. Hence the lemma is proved if we show that

$$\sup_{p\in\Lambda_1} P(c_N - a_N \leqq I(Z^{(N)}, p^0) \leqq c_N + a_N + I(Z^{(N)}, p) | p) \tag{2.24}$$

tends to zero for $N \to \infty$. By Lemma 2.2 it suffices to show that for every $A > 0$,

$$\sup_{p\in\Omega} P\left(c_N - a_N \leqq I(Z^{(N)}, p^0) \leqq c_N + a_N + \frac{A}{N} \middle| p\right) \to 0 \tag{2.25}$$

for $N \to \infty$. We consider three cases.

(i) Suppose that $p^0 \in \mathring{\Omega}$. Since $c_N + a_N + AN^{-1} \to 0$ for $N \to \infty$, there exists $\varepsilon > 0$ such that for all sufficiently large N the set $\{z | z \in \Omega, I(z, p^0) \leqq c_N + a_N + AN^{-1}\}$ is contained in the convex set $\{z | z \in \Omega, z_i \geqq \varepsilon \text{ for } i = 1, \cdots, k\}$. By Lemma 2.1 the supremum over Ω in (2.25) may therefore be replaced by the supremum over the set of all $p \in \Omega$ with $p \geqq \varepsilon$. Furthermore, we may again use the fact that under p

$$I(Z^{(N)}, p^0) = \sum_{i=1}^{k} Z_i^{(N)} \log \frac{p_i}{p_i^0} + I(Z^{(N)}, p) \qquad \text{a.s.} \tag{2.26}$$

and $0 \leqq I(Z^{(N)}, p) < \infty$ a.s. It follows from Lemma 2.2 that to prove (2.25) it is sufficient to show that for every $A > 0$ and $\varepsilon > 0$,

$$\sup_{p\geqq\varepsilon} P\left(c_N - a_N - \frac{A}{N} \leqq \sum_{i=1}^{k} Z_i^{(N)} \log \frac{p_i}{p_i^0} \leqq c_N + a_N + \frac{A}{N} \middle| p\right) \tag{2.27}$$

tends to zero for $N \to \infty$.

The condition $Nc_N \to \infty$ implies that c_N is positive for all sufficiently large N; together with the condition $Na_N^2 c_N^{-1} \to 0$ it also yields

$$a_N + \frac{A}{N} = o\left(\left(\frac{c_N}{N}\right)^{1/2}\right) = o(c_N) \tag{2.28}$$

for $N \to \infty$. Hence $c_N - a_N - AN^{-1} > 0$ for all sufficiently large N. As for $p = p^0$ the random variable in (2.27) is equal to 0 a.s., the supremum in (2.27) may be restricted to the set of all $p \neq p^0$ with $p \geqq \varepsilon$. Applying Lemma 2.3 we find that it suffices to show that for every $A > 0$ and $\varepsilon > 0$

$$\Phi\left(\frac{c_N + a_N + AN^{-1} - I(p, p^0)}{\sigma(p, p^0)} N^{1/2}\right) - \Phi\left(\frac{c_N - a_N - AN^{-1} - I(p, p^0)}{\sigma(p, p^0)} N^{1/2}\right) \tag{2.29}$$

tends to zero for $N \to \infty$, uniformly for all $p \neq p^0$ with $p \geqq \varepsilon$.

Define, for $N = 1, 2, \cdots$,

$$\begin{aligned} \Omega_{N,1} &= \left\{p \,\middle|\, p \in \Omega, p \neq p^0, p \geqq \varepsilon, I(p, p^0) \leqq \frac{c_N}{2}\right\}, \\ \Omega_{N,2} &= \left\{p \,\middle|\, p \in \Omega, p \neq p^0, p \geqq \varepsilon, I(p, p^0) > \frac{c_N}{2}\right\}. \end{aligned} \tag{2.30}$$

For $p \in \Omega_{N,1}$, (2.29) is bounded above by

$$1 - \Phi\left(\frac{\frac{1}{2}c_N - a_N - AN^{-1}}{\sigma(p, p^0)} N^{1/2}\right) \tag{2.31}$$

and by (2.28) and Lemma 2.4

$$\begin{aligned} \frac{\frac{1}{2}c_N - a_N - AN^{-1}}{\sigma(p, p^0)} N^{1/2} &\sim \frac{c_N N^{1/2}}{2\sigma(p, p^0)} \geqq \frac{c_N N^{1/2}}{2[M_2 I(p, p^0)]^{1/2}} \\ &\geqq \left(\frac{Nc_N}{2M_2}\right)^{1/2} \to \infty \text{ for } N \to \infty. \end{aligned} \tag{2.32}$$

For $p \in \Omega_{N,2}$, (2.29) is bounded above by

$$\begin{aligned} \frac{a_N + AN^{-1}}{\sigma(p, p^0)} N^{1/2} &\leqq (a_N + AN^{-1}) \left(\frac{N}{M_1 I(p, p^0)}\right)^{1/2} \\ &\leqq (a_N + AN^{-1}) \left(\frac{2N}{M_1 c_N}\right)^{1/2} \to 0 \end{aligned} \tag{2.33}$$

by the mean value theorem, Lemma 2.4 and (2.28). Hence the suprema of (2.29) over both $\Omega_{N,1}$ and $\Omega_{N,2}$ tend to zero which proves the lemma for $p^0 \in \mathring{\Omega}$.

(ii) Suppose that p^0 is a boundary point but not an extreme point of Ω; without loss of generality we assume that for some $2 \leqq m \leqq k - 1$, $p_i^0 \neq 0$ for

$i = 1, \cdots, m$ and $p_i^0 = 0$ for $i = m + 1, \cdots, k$. Since $I(z, p^0) = \infty$ if $z_i \neq 0$ for some $m + 1 \leqq i \leqq k$, the set $\{z | z \in \Omega, I(z, p^0) \leqq c_N + a_N + AN^{-1}\}$ is contained in the convex set $\{z | z \in \Omega, z_i = 0 \text{ for } i = m + 1, \cdots, k\}$. By Lemma 2.1 the supremum over Ω in (2.25) may therefore be replaced by the supremum over all $p \in \Omega$ with $p_i = 0$ for $i = m + 1, \cdots, k$. But under any p with $p_i = 0$ for $i = m + 1, \cdots, k$,

$$I(Z^{(N)}, p^0) = \sum_{i=1}^{m} Z_i^{(N)} \log \frac{Z_i^{(N)}}{p_i^0} \quad \text{a.s.} \tag{2.34}$$

and $(Z_1^{(N)}, \cdots, Z_m^{(N)})$ has a multinomial distribution with parameters N and $(p_1, \cdots, p_m)$. Thus we have reduced the problem of proving (2.25) to the same problem in a lower dimensional parameter space where $(p_1^0, \cdots, p_m^0)$ is now an interior point. This has been dealt with in (i).

(iii) Suppose that p^0 is an extreme point of Ω. This implies that $I(Z^{(N)}, p^0)$ can only assume the values 0 and ∞. Since $c_N - a_N > 0$ for all sufficiently large N, (2.25) is immediate. *Q.E.D.*

We remark that in the proof of Lemma 2.5 we have made use of the condition $c_N \to 0$ only to ensure that in case (i), for every $A > 0$

$$\{z | z \in \Omega, I(z, p^0) \leqq c_N + a_N + AN^{-1}\} \subset \{z | z \in \Omega, z \geqq \varepsilon\} \tag{2.35}$$

for some $\varepsilon > 0$ for all sufficiently large N, whereas in case (ii) it is needed that the same condition holds for the reduced lower dimensional problem. As $a_N + AN^{-1} = o(c_N)$ by (2.14), Lemma 2.5 will continue to hold if we replace the condition $c_N \to 0$ by the following assumption. For all sufficiently large N the set $\{z | z \in \Omega, I(z, p^0) \leqq c_N\}$ remains bounded away from the set of all points $z \in \Omega$ that have $z_i = 0$ for all i for which $p_i^0 = 0$ but also for at least one i with $p_i^0 \neq 0$. This extension of Lemma 2.5 is the main step in relaxing the condition $-\log \alpha_N = o(N)$ in Theorem 1.1 (see Section 3).

We complete this section by stating the result on large deviations of W. Hoeffding in [2] that we already referred to in Section 1. For a nonempty set $A \subset \Omega$ and $p \in \Omega$, define

$$I(A, p) = \inf_{z \in A} I(z, p) = \inf_{z \in A} \sum_{i=1}^{k} z_i \log \frac{z_i}{p_i}. \tag{2.36}$$

If A is empty we take $I(A, p) = +\infty$. We recall that for any $A \subset \Omega$, A^N denotes the set of all $z \in A$ for which Nz has integer coordinates.

LEMMA 2.6 (Hoeffding). *Uniformly for all $A \subset \Omega$ and all $p \in \Omega$,*

$$P(Z^{(N)} \in A | p) = \exp\{-NI(A^N, p) + O(\log N)\}. \tag{2.37}$$

Moreover, for any $p \in \Omega$ and any sequence $A_N \subset \Omega$ with complex complements,

$$I(A_N^N, p) = I(A_N, p) + O(N^{-1} \log N), \tag{2.38}$$

hence

$$P(Z^{(N)} \in A_N | p) = \exp\{-NI(A_N, p) + O(\log N)\}. \tag{2.39}$$

3. Proof of Theorem 1.1

The size α_N likelihood ratio test based on $Z^{(N)}$ for $H: p = p^0$ against $K: p \neq p^0$ rejects H if

$$I(Z^{(N)}, p^0) = \sum_{i=1}^{k} Z_i^{(N)} \log \frac{Z_i^{(N)}}{p_i^0} \geqq c_N \tag{3.1}$$

with possible randomization if equality occurs. For this case, where $\Lambda = \Omega$, Oosterhoff [3] showed that Theorem 1.1 holds under the additional assumptions that $p^0 \in \mathring{\Omega}$ and that α_N tends to zero slowly. In his proof he found that under his conditions $-\log \alpha_N \sim Nc_N$ for $N \to \infty$, which implies the conclusions concerning c_N in the following lemma.

LEMMA 3.1 (Oosterhoff). *Let p^0 be an arbitrary point of $\mathring{\Omega}$ and let R_N denote the shortcoming of the size α_N likelihood ratio test* (3.1) *for $H: p = p^0$ against $K: p \in \Omega - \{p^0\}$. If*

$$\lim_{N \to \infty} \alpha_N = 0, \qquad -\log \alpha_N = o(N^{1/6}) \qquad \text{for} \quad N \to \infty, \tag{3.2}$$

then

$$\lim_{N \to \infty} \sup_{p \neq p^0} R_N(p) = 0, \tag{3.3}$$

and $Nc_N \to \infty$, $c_N \to 0$ for $N \to \infty$.

We begin by removing, as far as possible, the restriction $p^0 \in \mathring{\Omega}$ in Lemma 3.1.

LEMMA 3.2. *Let p^0 be an arbitrary point of Ω and let R_N denote the shortcoming of the size α_N likelihood ratio test* (3.1) *for $H: p = p^0$ against $K: p \in \Omega - \{p^0\}$. If*

$$\lim_{N \to \infty} \alpha_N = 0, \qquad -\log \alpha_N = o(N^{1/6}) \qquad \text{for} \quad N \to \infty, \tag{3.4}$$

then

$$\lim_{N \to \infty} \sup_{p \neq p^0} R_N(p) = 0. \tag{3.5}$$

Moreover, $Nc_N \to \infty$, $c_N \to 0$ for $N \to \infty$ unless p^0 is an extreme point of Ω.

PROOF. If $p^0 \in \mathring{\Omega}$ Lemma 3.2 is merely a repetition of Lemma 3.1. If p^0 is an extreme point of Ω, then the likelihood ratio test is uniformly most powerful and hence its shortcoming is identically equal to zero for all N. We may therefore suppose that p^0 is a boundary point but not an extreme point of Ω; without loss of generality we assume that for some $2 \leqq m \leqq k - 1$, $p_i^0 \neq 0$ for $i = 1, \cdots, m$ and $p_i^0 = 0$ for $i = m + 1, \cdots, k$.

In this case any admissible size α_N test for $H: p = p^0$ against $K: p \neq p^0$ rejects H with probability one if $Z_i^{(N)} \neq 0$ for at least one $i = m + 1, \cdots, k$, and with probability $\phi_N(z_1, \cdots, z_m)$ if $Z_i^{(N)} = z_i$ for $i = 1, \cdots, m$ and $Z_i^{(N)} = 0$ for $i = m + 1, \cdots, k$. The size α_N likelihood ratio test (3.1) is of this type with

$$\phi_N(z_1, \cdots, z_m) = \begin{cases} 1 & \text{if} \quad \sum_{i=1}^m z_i \log \frac{z_i}{p_i^0} > c_N, \\ \delta & \text{if} \quad \sum_{i=1}^m z_i \log \frac{z_i}{p_i^0} = c_N, \\ 0 & \text{if} \quad \sum_{i=1}^m z_i \log \frac{z_i}{p_i^0} < c_N, \end{cases} \tag{3.6}$$

where $0 < \delta \leqq 1$.

Let us introduce an auxiliary random vector $\tilde{Z}^{(N)} = (\tilde{Z}_1^{(N)}, \cdots, \tilde{Z}_m^{(N)})$ such that $N\tilde{Z}^{(N)}$ has a multinomial distribution with parameters N and $\tilde{p} = (\tilde{p}_1, \cdots, \tilde{p}_m)$, where $\tilde{p}$ is any point in

$$\tilde{\Omega} = \{(y_1, \cdots, y_m) | \sum_{i=1}^m y_i = 1, y_i \geqq 0 \text{ for } i = 1, \cdots, m\}. \tag{3.7}$$

Since $P(Z_{m+1}^{(N)} = \cdots = Z_k^{(N)} = 0 | p^0) = 1$, we have for the size α_N likelihood ratio test as well as for any admissible size α_N test

$$\alpha_N = E(\phi_N(\tilde{Z}^{(N)}) | \tilde{p}^0), \tag{3.8}$$

where $\tilde{p} = (p_1^0, \cdots, p_m^0)$. For the power of such a test at $p \neq p^0$ we have

$$\beta_N(p) = \begin{cases} 1 & \text{if} \quad p_1 = \cdots = p_m = 0, \\ 1 - \pi^N + \pi^N E(\phi_N(\tilde{Z}^{(N)}) | \tilde{p}) & \text{otherwise}, \end{cases} \tag{3.9}$$

where

$$\pi = \sum_{i=1}^m p_i, \qquad \tilde{p}_i = \frac{p_i}{\pi} \qquad \text{for} \quad i = 1, \cdots, m. \tag{3.10}$$

For the random vector $\tilde{Z}^{(N)}$, consider the auxiliary problem of testing $\tilde{H} : \tilde{p} = \tilde{p}^0$ against $K : \tilde{p} \neq \tilde{p}^0$, where $\tilde{p}$ denotes the parameter vector of the distribution of $\tilde{Z}^{(N)}$. A test for this problem will reject $\tilde{H}$ with probability $\phi_N(z)$ if $\tilde{Z}^{(N)} = z$. Such a test has size α_N if and only if ϕ_N satisfies (3.8), and its power at $\tilde{p}$ is given by

$$\tilde{\beta}_N(\tilde{p}) = E(\phi_N(\tilde{Z}^{(N)}) | \tilde{p}). \tag{3.11}$$

Thus there exists a one to one correspondence between the class of size α_N tests for H based on $Z^{(N)}$ that reject H with probability one if $Z_i^{(N)} \neq 0$ for at least one $i = m + 1, \cdots, k$ and the class of all size α_N tests for $\tilde{H}$ based on $\tilde{Z}^{(N)}$. Here corresponding tests have the same function ϕ_N and hence by (3.9) and (3.11) we find that for all p with $p_i \neq 0$ for at least one $i = 1, \cdots, m$, their power functions satisfy

$$\beta_N(p) = 1 - \pi^N + \pi^N \tilde{\beta}_N(\tilde{p}), \tag{3.12}$$

where π and $\tilde{p}$ are defined by (3.10). Let β_N^+ and $\tilde{\beta}_N^+$ denote the size α_N envelope power functions for testing H on the basis of $Z^{(N)}$ and $\tilde{H}$ on the basis of $\tilde{Z}^{(N)}$, respectively. Since only admissible size α_N tests for H enter into the determination of β_N^+, it follows from (3.9) and (3.12) that

$$\beta_N^+(p) = \begin{cases} 1 & \text{if } p_1 = \cdots = p_m = 0, \\ 1 - \pi^N + \pi^N \tilde{\beta}_N^+(\tilde{p}) & \text{otherwise,} \end{cases} \tag{3.13}$$

where π and $\tilde{p}$ are defined by (3.10).

The likelihood ratio test for the auxiliary problem of testing $\tilde{H}$ against $\tilde{K}$ is based on the statistic $I(\tilde{Z}^{(N)}, \tilde{p}^0)$. As the function ϕ_N for the size α_N likelihood ratio test given by (3.6) satisfies (3.8), this function is also the test function of the size α_N likelihood ratio test for $\tilde{H}$ against $\tilde{K}$. In the first place this implies that the critical values of the two size α_N likelihood ratio tests are both equal to the same number c_N. In the second place it means that (3.13) will continue to hold if the envelope power functions β_N^+ and $\tilde{\beta}_N^+$ are replaced by the power functions β_N and $\tilde{\beta}_N$ of the size α_N likelihood ratio tests. Hence, if R_N and $\tilde{R}_N$ denote the shortcomings of the size α_N likelihood ratio tests for H against K and for $\tilde{H}$ against $\tilde{K}$, respectively, then

$$R_N(p) = \begin{cases} 0 & \text{if } p_1 = \cdots = p_m = 0, \\ \pi^N \tilde{R}_N(\tilde{p}) & \text{otherwise,} \end{cases} \tag{3.14}$$

where π and $\tilde{p}$ are defined by (3.10). Since $\pi \leqq 1$ and $\tilde{R}_N(\tilde{p}^0) = 0$,

$$\sup_{p \neq p^0} R_N(p) \leqq \sup_{\tilde{p} \neq \tilde{p}^0} \tilde{R}_N(\tilde{p}). \tag{3.15}$$

As $\tilde{p}^0$ is an interior point of $\tilde{\Omega}$ we may apply Lemma 3.1 to the auxiliary testing problem to conclude that the right side of (3.15) tends to zero and that $Nc_N \to \infty$, $c_N \to 0$ for $N \to \infty$. *Q.E.D.*

Our next step will be to remove the restriction $\Lambda = \Omega$.

LEMMA 3.3. *Let Λ be an arbitrary subset of Ω, p^0 an arbitrary point of Λ and let R_N denote the shortcoming of the size α_N likelihood ratio test based on $Z^{(N)}$ for $H: p = p^0$ against $K: p \in \Lambda_1 = \Lambda - \{p^0\}$. If*

$$\lim_{N \to \infty} \alpha_N = 0, \qquad -\log \alpha_N = o(N^{1/6}) \qquad \text{for} \quad N \to \infty, \tag{3.16}$$

then

$$\lim_{N \to \infty} \sup_{p \in \Lambda_1} R_N(p) = 0. \tag{3.17}$$

PROOF. If p^0 is an extreme point of Ω, the likelihood ratio test for H against K is uniformly most powerful against K and hence its shortcoming is equal to zero for all $p \in \Lambda_1$ and all N. We may therefore suppose that p is not an extreme point of Ω.

The size α_N likelihood ratio test for H against K rejects H if

$$\sup_{\pi \in \Lambda} \sum_{i=1}^{k} Z_i^{(N)} \log \frac{\pi_i}{p_i^0} \geqq c_N^*, \tag{3.18}$$

possibly with randomization if equality occurs. Let us compare this test with the size α_N likelihood ratio test (3.1) for H against $p \neq p^0$. By Lemma 3.2 the shortcoming of the latter test vanishes uniformly for all $p \neq p^0$ for $N \to \infty$ and hence Lemma 3.3 will be proved if we show that

$$\sup_{p\in\Lambda_1} P\left(\sup_{\pi\in\Lambda} \sum_{i=1}^{k} Z_i^{(N)} \log \frac{\pi_i}{p_i^0} \leqq c_N^*, I(Z^{(N)}, p^0) \geqq c_N \middle| p\right) \tag{3.19}$$

tends to zero for $N \to \infty$, where c_N is the constant that occurs in (3.1). As p^0 is not an extreme point of Ω, Lemma 3.2 also ensures that $c_N \to 0$ and $Nc_N \to \infty$ for $N \to \infty$. Furthermore we note that under any $p \in \Lambda$

$$\sup_{\pi\in\Lambda} \sum_{i=1}^{k} Z_i^{(N)} \log \frac{\pi_i}{p_i^0} \leqq I(Z^{(N)}, p^0) \qquad \text{a.s.} \tag{3.20}$$

Since the tests (3.1) and (3.18) have the same size it follows that c_N^* and c_N may be chosen in such a way that $c_N^* \leqq c_N$. To prove Lemma 3.3 it is therefore sufficient to show that

$$\sup_{p\in\Lambda_1} P\left(\sup_{\pi\in\Lambda} \sum_{i=1}^{k} Z_i^{(N)} \log \frac{\pi_i}{p_i^0} \leqq c_N, I(Z^{(N)}, p^0) \geqq c_N \middle| p\right) \tag{3.21}$$

tends to zero for $N \to \infty$. As $c_N \to 0$ and $Nc_N \to \infty$ for $N \to \infty$, this is the content of Lemma 2.5 for $a_N = 0$. *Q.E.D.*

We now turn to the case where α_N tends to zero fast.

LEMMA 3.4. *Lemma* 3.3 *holds if the conditions* (3.16) *concerning* α_N *are replaced by*

$$\lim_{N\to\infty} \frac{-\log \alpha_N}{(\log N)^2} = \infty, \qquad -\log \alpha_N = o(N) \qquad \text{for} \quad N \to \infty. \tag{3.22}$$

PROOF. For the same reason as in the proof of Lemma 3.3 we may restrict attention to the case where p^0 is not an extreme point of Ω. Consider the size α_N likelihood ratio test (3.1) for $H: p = p^0$ against $p \neq p^0$. The convexity of $I(z, p^0)$ in z ensures that the sets

$$\begin{aligned} A_N &= \{z \mid z \in \Omega, I(z, p^0) \geqq c_N\}, \\ B_N &= \{z \mid z \in \Omega, I(z, p^0) > c_N\} \end{aligned} \tag{3.23}$$

have convex complements. By the second part of Lemma 2.6

$$\begin{aligned} \alpha_N &\leqq P(I(Z^{(N)}, p^0) \geqq c_N | p^0) \\ &= \exp\{-NI(A_N, p^0) + O(\log N)\} = \exp\{-Nc_N + O(\log N)\}, \end{aligned} \tag{3.24}$$

or $Nc_N \leqq -\log \alpha_N + O(\log N)$. This implies that $c_N \to 0$ for $N \to \infty$ by the second part of (3.22). For $z \in \Omega$, the function $I(z, p^0)$ assumes all values in the interval $[0, -\log p_m^0]$ where p_m^0 is the smallest positive coordinate of p^0. As p^0 is not an extreme point of Ω, $-\log p_m^0 > 0$ and hence $0 \leqq c_N < -\log p_m^0$ for all sufficiently large N. For these values of N, $I(B_N, p^0) = c_N$ and by the second part of Lemma 2.6

$$\alpha_N \geqq P(I(Z^{(N)}, p^0) > c_N|p^0) = \exp\{-Nc_N + O(\log N)\}. \tag{3.25}$$

Hence

$$\alpha_N = \exp\{-Nc_N + O(\log N)\}, \qquad Nc_N = -\log \alpha_N + O(\log N); \tag{3.26}$$

together with (3.22) this yields

$$\lim_{N\to\infty} \frac{Nc_N}{(\log N)^2} = \infty, \qquad \lim_{N\to\infty} c_N = 0. \tag{3.27}$$

By the first part of Lemma 2.6 there exists a number $0 \leqq a < \infty$ independent of N, such that for every N and every $z^{(N)} \in \Omega^N$ with $I(z^{(N)}, p^0) < c_N - a(\log N)/N$,

$$P(Z^{(N)} = z^{(N)}|p^0) \geqq \exp\{-Nc_N + a\log N + O(\log N)\} \geqq N\alpha_N. \tag{3.28}$$

Obviously, any size α_N test for $H: p = p^0$ cannot reject H with probability larger than N^{-1} if $Z^{(N)}$ assumes one of these values $z^{(N)}$. Hence the size α_N envelope power β_N^+ for testing H satisfies

$$\beta_N^+(p) \leqq N^{-1} + P(I(Z^{(N)}, p^0) \geqq c_N - a_N|p) \tag{3.29}$$

for all $p \neq p^0$, where

$$a_N = \frac{a\log N}{N}, \qquad 0 \leqq a < \infty. \tag{3.30}$$

We note that (3.29) is a slightly modified form of a conclusion due to W. Hoeffding in [2].

It follows from (3.29) that the shortcoming $R_N(p)$ at p of the size α_N likelihood ratio test (3.18) for H against K is bounded above by

$$P\left(\sup_{\pi\in\Lambda} \sum_{i=1}^{k} Z_i^{(N)} \log \frac{\pi_i}{p_i^0} \leqq c_N^*, I(Z^{(N)}, p^0) \geqq c_N - a_N \Big| p\right) + \frac{1}{N}. \tag{3.31}$$

By the reasoning given in the proof of Lemma 3.3 we may assume that $c_N^* \leqq c_N$ and hence Lemma 3.4 is proved if we show that

$$\sup_{p\in\Lambda_1} P\left(\sup_{\pi\in\Lambda} \sum_{i=1}^{k} Z_i^{(N)} \log \frac{\pi_i}{p_i^0} \leqq c_N, I(Z^{(N)}, p^0) \geqq c_N - a_N \Big| p\right) \tag{3.32}$$

tends to zero for $N \to \infty$. By (3.27) and (3.30), $c_N \to 0$, $Nc_N \to \infty$ and $Na_N^2/c_N \to 0$ for $N \to \infty$. Application of Lemma 2.5 completes the proof.

Proof of Theorem 1.1. The theorem is proved by splitting up the sequence α_N into two subsequences satisfying (3.16) and (3.22), respectively, and applying Lemmas 3.3 and 3.4.

In Section 1 we claimed that the condition $-\log \alpha_N = o(N)$ in Theorem 1.1 may be relaxed. To see how this can be achieved we obviously need not consider the proof of Theorem 1.1 for the case where $\alpha_N \to 0$ slowly; we only have to inspect the proof of Lemma 3.4.

In proving Lemma 3.4 we have made use of the condition $-\log \alpha_N = o(N)$ only to conclude that $c_N \to 0$ for $N \to \infty$, provided that p^0 is not an extreme point of Ω. This fact was needed on two occasions. In the first place it was used to ensure that, if p^0 is not an extreme point of Ω, we have $0 \leqq c_N < -\log p_m^0$ for all sufficiently large N, where p_m^0 denotes the smallest positive coordinate of p^0. As the function $I(z, p^0)$ assumes its largest finite value $-\log p_m^0$ at those extreme points $z \in \Omega$ for which $z_i = 1$ for some i with $p_i^0 = p_m^0$, the assertion $0 \leqq c_N < -\log p_m^0$ is equivalent to saying that the set

$$C_N = \{z \mid z \in \Omega, I(z, p^0) \leqq c_N\} \tag{3.33}$$

does not contain these specific extreme points of Ω. We recall that C_N is the closure of the acceptance region of the size α_N likelihood ratio test (3.1) for $H: p = p^0$ against $p \neq p^0$.

In the second place, the fact that $c_N \to 0$ was used to ensure applicability of Lemma 2.5. However, in the remark following the proof of this lemma we pointed out that the lemma remains valid if the condition $c_N \to 0$ is replaced by the following assumption.

ASSUMPTION 1. *For all sufficiently large N the sets C_N defined in* (3.33) *remain bounded away from the set D_{p^0} of all points $z \in \Omega$ that have $z_i = 0$ for all i for which $p_i^0 = 0$ but also for at least one i with $p_i^0 \neq 0$.*

This assumption obviously implies that, for all sufficiently large N, the set C_N does not contain any extreme points of Ω, unless p^0 itself is an extreme point. It follows that Theorem 1.1 will continue to hold if the condition $-\log \alpha_N = o(N)$ is replaced by Assumption 1. Note that Assumption 1 imposes no restriction if p^0 is an extreme point of Ω.

One easily verifies that for $p^0 < 1$ (that is, $p_i^0 < 1$ for all i),

$$\inf_{z \in D_{p^0}} I(z, p^0) = -\log (1 - p_m^0), \tag{3.34}$$

where p_m^0 is defined as above. Since $I(z, p^0)$ is convex and uniformly continuous on the set of all z that have $z_i = 0$ for all i with $p_i^0 = 0$, Assumption 1 is equivalent to the requirement that if $p^0 < 1$, there exists $\varepsilon > 0$ such that for all sufficiently large N, $c_N \leqq -\log (1 - p_m^0) - \varepsilon$. Going over the proof of Lemma 3.4 we find that this, in turn, is equivalent to

ASSUMPTION 2. *There exists $\varepsilon > 0$ such that for all sufficiently large N, $-\log \alpha_N \leqq N\big(-\log (1 - p_m^0) - \varepsilon\big)$, where p_m^0 denotes the smallest positive coordinate of p^0.*

Note that if p^0 is an extreme point of Ω, Assumption 2 imposes no restriction on the sequence α_N. As Assumptions 1 and 2 are equivalent, the condition $-\log \alpha_N = o(N)$ in Theorem 1.1 may be replaced by the obviously weaker Assumption 2.

By sharpening Lemmas 2.3 and 2.5 one can show that Theorem 1.1 will still continue to hold if C_N does approach D_{p^0} for $N \to \infty$, but does so sufficiently slowly. In Assumption 2 this corresponds to allowing ε to tend to zero for $N \to \infty$, provided that this convergence is sufficiently slow.

4. The case of a composite hypothesis

In this section we show by means of a counterexample that Theorem 1.1 breaks down in the case of a composite hypothesis H even when $\Lambda = \Omega$. We consider the binomial case $k = 2$ and write $Z^{(N)} = Z_1^{(N)}$, $1 - Z^{(N)} = Z_2^{(N)}$, $p = p_1$ and $1 - p = p_2$. Thus $NZ^{(N)}$ has a binomial distribution with parameters N and p where p is an arbitrary point in $[0, 1]$. For $z \in [0, 1]$, $p \in [0, 1]$ and $\Lambda_0 \subset [0, 1]$ we define

$$I(z, p) = z \log \frac{z}{p} + (1 - z) \log \frac{1 - z}{1 - p}. \tag{4.1}$$

and

$$I(z, \Lambda_0) = \inf_{p \in \Lambda_0} I(z, p). \tag{4.2}$$

If Λ_0 is a proper subset of $[0, 1]$, one may consider the problem of testing $H: p \in \Lambda_0$ against $K: p \notin \Lambda_0$. A nonrandomized likelihood ratio test for H against K rejects H if

$$I(Z^{(N)}, \Lambda_0) \geqq \tilde{c}_N; \tag{4.3}$$

the size of this test is

$$\alpha_N = \sup_{p \in \Lambda_0} P(I(Z^{(N)}, \Lambda_0) \geqq \tilde{c}_N | p). \tag{4.4}$$

Consider any fixed sequence of positive numbers $\tilde{c}_N$ such that

$$\lim_{N \to \infty} \tilde{c}_N = 0, \qquad \lim_{N \to \infty} N\tilde{c}_N = \infty. \tag{4.5}$$

We choose two positive integers a and b and a sequence d_N such that $0 < d_N < \tilde{c}_N$ for all N and $Nd_N \to 0$ for $N \to \infty$. Next we construct a set $\Lambda_0 \subset [0, 1]$ with the following property: there exists an infinite sequence of positive integers $N_1 < N_2 < \cdots$ such that for every i the following conditions are satisfied:

(i) Λ_0 contains points $p_{i,1} < p_{i,2}$ with

$$I\left(\frac{a}{N_i}, p_{i,j}\right) = \tilde{c}_{N_i} - d_{N_i} \qquad \text{for } j = 1, 2. \tag{4.6}$$

(ii) Λ_0 contains points $p_{i,3} < p_{i,4}$ with

$$I\left(1 - \frac{b}{N_i}, p_{i,j}\right) = \tilde{c}_{N_i} \qquad \text{for } j = 3, 4. \tag{4.7}$$

(iii) Λ_0 does not contain points in $(p_{i,1}, p_{i,2}) \cup (p_{i,3}, p_{i,4})$.

To see that this construction is possible we note that for sufficiently large N, $1 \leqq a, b \leqq N - 1$ and hence

$$(4.8) \qquad I\left(\frac{a}{N}, 0\right) = I\left(\frac{a}{N}, 1\right) = I\left(1 - \frac{b}{N}, 0\right) = I\left(1 - \frac{b}{N}, 1\right) = \infty.$$

Thus, for any sequence $N_1 < N_2 < \cdots$ with $N_1 - 1 \geqq \max(a, b)$, points $p_{i,j}$ with properties (i) and (ii) exist for every i. Notice that obviously $0 < p_{i,1} < aN_i^{-1} < p_{i,2}$ and $p_{i,3} < 1 - bN_i^{-1} < p_{i,4} < 1$. Since $\tilde{c}_N \to 0$ for $N \to \infty$, we can also ensure for every $0 < \varepsilon < \frac{1}{2}$ that $p_{1,2} < \varepsilon < 1 - \varepsilon < p_{1,3}$ by choosing N_1 large enough. Having chosen $N_1 \geqq \max(a, b) + 1$ in such a way that the above holds for some $0 < \varepsilon < \frac{1}{2}$, we proceed to choose N_i for $i = 2, 3, \cdots$ sequentially in such a way that

$$(4.9) \qquad \begin{aligned} &\frac{a}{N_i} < p_{i-1,1}, \qquad I\left(\frac{a}{N_i}, p_{i-1,1}\right) > \tilde{c}_{N_i} - d_{N_i}, \\ &1 - \frac{b}{N_i} > p_{i-1,4}, \qquad I\left(1 - \frac{b}{N_i}, p_{i-1,4}\right) > \tilde{c}_{N_i}. \end{aligned}$$

This is clearly possible as $p_{i-1,1} > 0, I(0, p_{i-1,1}) > 0, p_{i-1,4} < 1, I(1, p_{i-1,4}) > 0$ for all $i \geqq 2$ and $\tilde{c}_N \to 0$ for $N \to \infty$. However, this implies that $p_{i,2} < p_{i-1,1}$ and $p_{i,3} > p_{i-1,4}$ for every $i \geqq 2$. Because we already made sure that $p_{1,2} < \varepsilon < 1 - \varepsilon < p_{1,3}$, condition (iii) will be satisfied if Λ_0 does not contain other points in an ε neighborhood of 0 and 1 besides the points $p_{i,j}$.

For an arbitrary sequence $\tilde{c}_N$ satisfying (4.5) and for a corresponding set Λ_0 that we have just constructed, we consider the sequence of likelihood ratio tests (4.3) for $H: p \in \Lambda_0$ against $K: p \notin \Lambda_0$. We shall show that α_N defined by (4.4) satisfies the conditions $\alpha_N \to 0$ and $-\log \alpha_N = o(N)$ of Theorem 1.1, but that the shortcoming of this sequence of likelihood ratio tests does not tend to zero uniformly for all $p \notin \Lambda_0$.

By (4.2) and (4.4)

$$(4.10) \qquad \alpha_N \leqq \sup_{p \in \Lambda_0} P(I(Z^{(N)}, p) \geqq \tilde{c}_N | p),$$

and since $N\tilde{c}_N \to \infty$, $\alpha_N \to 0$ for $N \to \infty$ by Lemma 2.2. Let p_0 be an isolated point of Λ_0 with $0 < p_0 < 1$, for example, $p_0 = p_{1,1}$. For z in a sufficiently small neighborhood of p_0, $I(z, \Lambda_0) = I(z, p_0)$ and the absolute value of the derivative of this function is smaller than δ. Since $\tilde{c}_N \to 0$, the set

$$(4.11) \qquad \tilde{A}_N = \{z | 0 \leqq z \leqq 1, I(z, \Lambda_0) \geqq \tilde{c}_N\}$$

will contain, for all sufficiently large N, a point $z^{(N)}$ for which $Nz^{(N)}$ is an integer and $I(z^{(N)}, p_0) \leqq \tilde{c}_N + \delta N^{-1}$. Hence by Lemma 2.6

$$(4.12) \qquad \alpha_N \geqq P(I(Z^{(N)}, \Lambda_0) \geqq \tilde{c}_N | p_0) \geqq \exp\{-N\tilde{c}_N - \delta + O(\log N)\},$$

and as $\tilde{c}_N \to 0$, $-\log \alpha_N = o(N)$ for $N \to \infty$.

For $N = N_i$ we need a sharper asymptotic lower bound for α_N. By properties (ii) and (iii) of the set Λ_0

$$I\left(1 - \frac{b}{N_i}, \Lambda_0\right) = I\left(1 - \frac{b}{N_i}, p_{i,3}\right) = \tilde{c}_{N_i} \tag{4.13}$$

for all i. It follows that for every $\varepsilon > 0$ we have for all sufficiently large i

$$\begin{aligned} \alpha_{N_i} &\geqq P\left(Z^{(N_i)} = 1 - \frac{b}{N_i} | p_{i,3}\right) \\ &= \exp\{-N_i \tilde{c}_{N_i}\} P\left(Z^{(N_i)} = 1 - \frac{b}{N_i} | 1 - \frac{b}{N_i}\right) \\ &\geqq (1 - \varepsilon) e^{-b} \frac{b^b}{b!} \exp\{-N_i \tilde{c}_{N_i}\}. \end{aligned} \tag{4.14}$$

Also, by properties (i) and (iii) of the set Λ_0

$$I\left(\frac{a}{N_i}, \Lambda_0\right) = I\left(\frac{a}{N_i}, p_{i,j}\right) = \tilde{c}_{N_i} - d_{N_i} \tag{4.15}$$

for $j = 1, 2$ and all i. Because $N d_N \to 0$ for $N \to \infty$ this implies that for every $\varepsilon > 0$

$$\begin{aligned} \sup_{p \in \Lambda_0} P\left(Z^{(N_i)} = \frac{a}{N_i} | p\right) &= \max_{j=1,2} P\left(Z^{(N_i)} = \frac{a}{N_i} | p_{i,j}\right) \\ &= \exp\{-N_i(\tilde{c}_{N_i} - d_{N_i})\} P\left(Z^{(N_i)} = \frac{a}{N_i} | \frac{a}{N_i}\right) \\ &\leqq (1 + \varepsilon) e^{-a} \frac{a^a}{a!} \exp\{-N_i \tilde{c}_{N_i}\} \end{aligned} \tag{4.16}$$

for all sufficiently large i. Together with (4.14) this implies that there exists a number $0 < \phi \leqq 1$ such that the test T_N that rejects H with probability ϕ if $Z^{(N)} = aN^{-1}$ has size at most α_N whenever $N = N_i$ and i is sufficiently large. Hence, if β_N^+ denotes the size α_N envelope power for testing H, we have shown that for every $\varepsilon > 0$

$$\beta_{N_i}^+ \left(\frac{a}{N_i}\right) \geqq \phi P\left(Z^{(N_i)} = \frac{a}{N_i} | \frac{a}{N_i}\right) \geqq \phi(1 - \varepsilon) e^{-a} \frac{a^a}{a!}, \tag{4.17}$$

for all sufficiently large i. On the other hand, property (i) of the set Λ_0 ensures that for $N = N_i$ the critical region $\tilde{A}_N$ of the likelihood ratio test does not contain points in the interval $[p_{i,1}, p_{i,2}]$. If β_N denotes the power of the size α_N likelihood ratio test, this means that for all i

$$\beta_{N_i}\left(\frac{a}{N_i}\right) \leqq \beta_{N_i}(p_{i,1}) + \beta_{N_i}(p_{i,2}) \leqq 2\alpha_{N_i}, \tag{4.18}$$

where the right side tends to zero for $i \to \infty$. Together with (4.17) this proves that the shortcoming of the likelihood ratio test does not tend to zero uniformly for all $p \notin \Lambda_0$.

REFERENCES

[1] W. Feller, *An Introduction to Probability Theory and Its Applications*, Vol. 2, New York, Wiley, 1966.

[2] W. Hoeffding, "Asymptotically optimal tests for multinomial distributions," *Ann. Math. Statist.*, Vol. 36 (1965), pp. 369–401.

[3] J. Oosterhoff, *Combination of One-sided Statistical Tests*, Amsterdam, Mathematisch Centrum, 1969.