The Annals of Statistics
1981, Vol. 9, No. 6, 1248–1257

LIKELIHOOD RATIO TESTS FOR AND AGAINST A STOCHASTIC ORDERING BETWEEN MULTINOMIAL POPULATIONS[1]

By Tim Robertson and F. T. Wright

University of Iowa and University of Missouri-Rolla

Likelihood ratio tests concerning the parameters of two multinomial populations are discussed. A stochastic ordering restriction is considered as a one sided alternative to equality. The one and two sample tests for equality versus stochastic ordering and stochastic ordering versus all alternatives are derived and their large sample distributions are obtained. The large sample distributions are mixtures of chi-squared distributions. The tests developed provide discrete analogues for the one sided Mann-Whitney-Wilcoxon and Kolmogorov-Smirnov tests.

1. Introduction. Tests for the equality of two populations against a stochastically ordered alternative are among the more widely used nonparametric procedures. They include the one-sided Mann-Whitney-Wilcoxon and Kolmogorov-Smirnov tests. We consider analogous one- and two-sample likelihood ratio procedures under the assumption that the underlying populations are discrete. It is well known that one-sided procedures are more powerful than their two-sided counterparts. Thus these procedures are recommended over the standard chi-squared tests provided, of course, that the underlying assumptions are valid.

We denote the two collections of multinomial parameters by $p = (p_1, p_2, \cdots, p_k)$ and $q = (q_1, q_2, \cdots, q_k)$ and we assume that both p and q are in $A = \{(x_1, x_2, \cdots, x_k): x_i > 0, \sum_{i=1}^k x_i = 1\}$. Consider the hypothesis H_1 that the q distribution is stochastically larger than the p distribution. Specifically,

$$\text{(1.1)} \qquad H_1: \sum_{j=1}^i p_j \geq \sum_{j=1}^i q_j, \qquad i = 1, 2, \cdots, k-1, \quad \sum_{j=1}^k p_j = \sum_{j=1}^k q_j.$$

We note that the restrictions in (1.1) determine a partial order on R^n, which we denote symbolically by $p \gg q$. Robertson and Wright (1980b) studied this partial order as a quantification of the notion of conformity to an order restriction. In particular, one interpretation of $p \gg q$ is that p satisfies the hypothesis that its entries are decreasing more than does q. In this paper, we consider likelihood ratio statistics for testing problems involving three hypotheses, namely, $H_0: p = q$, H_1 and $H_2 = \sim (\text{not } H_1)$. We shall consider both one- and two-sample tests.

Chacko (1966) studied a likelihood ratio statistic for testing the null hypothesis that $p = q_0 = k^{-1}(1, 1, \cdots, 1)$ against the alternative that $p_1 \geq p_2 \geq \cdots \geq p_k$ (and of course, $p \neq q_0$). The hypothesis $p \gg q_0$ is implied by the hypothesis $p_1 \geq p_2 \geq \cdots \geq p_k$, but not conversely, so that the test discussed here has a less restrictive alternative than the one considered by Chacko. In fact, $p \gg q_0$ is equivalent to $i^{-1} \sum_{j=1}^i p_j \geq (k - i)^{-1} \sum_{j=i+1}^k p_j$ which, for lack of a better phrase, we term "decreasing on the average". Thus, the results in this paper yield maximum likelihood estimates under this restriction and distribution theory for likelihood ratio tests of homogeneity versus decreasing on the average and for testing decreasing on the average as a null hypothesis.

It is interesting to note that the statistic, derived in Section 3, for testing $p = q_0$ versus

---

Received August 1980; revised April 6, 1981.

[1] This research was partially sponsored by the office of Naval Research under Contracts ONR N00014-80-C0321 and ONR N00014-80-C-0322.

AMS 1970 *subject classification.* Primary 62G10; secondary 62E20.

Key words and phrases. Stochastic ordering, multinomial distributions, chi-bar-squared tests, likelihood ratio, asymptotic results.

$p \gg q_0$ has a chi-bar-squared distribution as did Chacko's test statistic. Robertson (1978) generalized Chacko's work by considering the test of $p = q$ (arbitrary q) against an arbitrary order restriction on p. He also considered the problem of testing an order restriction on p as a null hypothesis. The two-sample test developed here for H_0 versus $H_1 - H_0$ is discussed in Grove (1980).

In Section 2, the one- and two-sample maximum likelihood estimates of the multinomial parameters subject to the restrictions in H_1 are derived. The distribution theory for the one-sample tests of H_0 versus $H_1 - H_0$ and H_1 versus H_2 is given in Section 3 and Section 4 contains the corresponding two-sample theory.

2. Restricted maximum likelihood estimates. In order to develop the desired likelihood ratio tests we must first obtain the maximum likelihood estimates under the restriction $p \gg q$. The approach uses the theory given in Section 5 of Barlow and Brunk (1972) which requires the following notation. For any collection of positive weights, $w = (w_1, w_2, \cdots, w_k)$, let $(x, y)_w$ be the inner product on R^k defined by $(x, y)_w = \sum_{i=1}^k x_i y_i w_i$; let $\|\cdot\|_w$ denote the induced norm, i.e., $\|x\|_w^2 = \sum_{i=1}^k x_i^2 w_i$; and for any subset A of R^k let $E_w(x \mid A)$ denote the projection (i.e., closest point under $\|\cdot\|_w$) of x onto A provided it exists and is unique (cf. Brunk, 1965).

We first consider the one-sample problem. Assume q is known; assume a random sample of size m from the population associated with p and let $\hat{p} = (\hat{p}_1, \hat{p}_2, \cdots, \hat{p}_k)$ be the vector of relative frequencies; i.e., $m\hat{p}$ has a multinomial distribution with parameters m and p. Let $C = \{x \in R^k: x_1 \geq x_2 \geq \cdots \geq x_k\}$ and note that C is a closed convex cone in R^k, so that by Brunk (1965), $E_w(\cdot \mid C)$ is well defined.

THEOREM 2.1. *If $\hat{p}_i > 0$, $i = 1, 2, \cdots, k$, then the maximum likelihood estimate (m.l.e.) of p subject to H_1 is given by*

$$\bar{p} = \hat{p} E_{\hat{p}}(q/\hat{p} \mid C), \tag{2.1}$$

where, for $x, y \in R^k$, xy denotes the vector $(x_1 y_1, x_2 y_2, \cdots, x_k y_k)$ and $x/y = (x_1/y_1, x_2/y_2, \cdots, x_k/y_k)$.

Before the proof of Theorem 2.1 is given, we describe the lower sets algorithm (LSA) for computing $E_w(x \mid C)$. For A a nonempty subset of $\{1, 2, \cdots, k\}$, set

$$M(A) = \textstyle\sum_{i \in A} w_i x_i / \sum_{i \in A} w_i.$$

Set $i_0 = 0$ and choose i_1 the largest positive integer i which maximizes $M(\{i_0 + 1, \cdots, i\})$. Next choose i_2 the largest integer i greater than i_1 which maximizes $M(\{i_1 + 1, \cdots, i\})$. Continuing this process, we obtain $0 = i_0 < i_1 < \cdots < i_\ell = k$ and the projection

$$E_w(x \mid C)_i = M(\{i_{j-1} + 1, \cdots, i_j\}) \quad \text{for } i \in \{i_{j-1} + 1, \cdots, i_j\} \quad \text{and} \quad j = 1, 2, \cdots, \ell.$$

The sets $\{i_{j-1} + 1, \cdots, i_j\}$ are called the level sets.

PROOF. The m.l.e., $\bar{p} = (\bar{p}_1, \bar{p}_2, \cdots, \bar{p}_k)$, solves the following optimization problem:

$$\text{minimize} \quad -\textstyle\sum_{i=1}^k m\hat{p}_i \ln p_i \quad \text{subject to} \quad p \gg q.$$

Set $s = m^{-1}(p_1/\hat{p}_1, p_2/\hat{p}_2, \cdots, p_k/\hat{p}_k)$, $w = m(\hat{p}_1, \hat{p}_2, \cdots, \hat{p}_k)$, $g = m^{-1}(q_1/\hat{p}_1, q_2/\hat{p}_2, \cdots, q_k/\hat{p}_k)$, and $\phi(y) = -\ln y$. Then $\bar{s} = m^{-1}(\bar{p}_1/\hat{p}_1, \bar{p}_2/\hat{p}_2, \cdots, \bar{p}_k/\hat{p}_k)$ solves

$$\text{minimize} \textstyle\sum_{i=1}^k w_i \phi(s_i) \tag{2.2}$$

subject to

$$\textstyle\sum_{j=1}^i w_j(g_j - s_j) \leq 0, \; 1 \leq i < k \quad \text{and} \quad \sum_{j=1}^k w_j(g_j - s_j) = 0.$$

The Fenchel dual, C^{w*}, of C is

$$C^{w*} = \{u;\ (u, v)_w \le 0 \text{ for each } v \in C\}$$
$$= \{u;\ \textstyle\sum_{j=1}^{i} u_j w_j \le 0,\ 1 \le i < k,\ \sum_{j=1}^{k} u_j w_j = 0\}$$

(cf. Barlow *et al.*, 1972, page 49). Thus (2.2) becomes

$$\text{minimize } \textstyle\sum_{i=1}^{k} w_i \phi(s_i) \quad \text{subject to} \quad g - s \in C^{w*}$$

and by Theorem 3.4 of Barlow and Brunk (1972) the solution to (2.2) is unique and is the projection of g onto the cone C. Thus

$$\bar{p} = m\hat{p} E_w(q/m\hat{p} \mid C) = \hat{p} E_{\hat{p}}(q/\hat{p} \mid C).$$

THEOREM 2.2. *As $m \to \infty$, $\bar{p}$ converges almost surely to p provided $p \gg q$.*

PROOF. By the strong law of large numbers, $\hat{p} \to p$ a.s. as $m \to \infty$. Moreover, $E_w(x \mid C)$ is continuous in both w and x so that $\bar{p} \to pE_p(q/p \mid C)$ a.s. Using the LSA to compute $E_p(q/p \mid C)$, one sees that since $p \gg q$, $M(\{1, \cdots, i\}) \le 1$ with equality for $i = k$. Hence, $E_p(q/p) \mid C) = e_k$ where e_k is the k-dimensional vector of ones and so $pE_p(q/p) \mid C) = p$.

In the two-sample problem let $\hat{q}$ denote the vector of relative frequencies of a sample of size n from the q population and assume that $\hat{p}$ and $\hat{q}$ are independent. Let

$$B = \{x \in R^{2k};\ x_1 \ge x_2 \ge \cdots \ge x_k,\ x_{k+1} \le x_{k+2} \le \cdots \le x_{2k}\},$$
$$N = m + n \quad \text{and} \quad \theta = (p, q).$$

THEOREM 2.3. *If $\hat{p}_i, \hat{q}_i > 0$, $i = 1, 2, \cdots, k$, then the maximum likelihood estimate of θ subject to H_1 is given by*

$$(\bar{p}, \bar{q}) = \bar{\theta} = wE_w(h \mid B) \tag{2.3}$$

where $w = (m\hat{p}_1, m\hat{p}_2, \cdots, m\hat{p}_k, n\hat{q}_1, n\hat{q}_2, \cdots, n\hat{q}_k)$ and

$$h_i = \begin{cases} N^{-1} + \dfrac{n}{mN}\dfrac{\hat{q}_i}{\hat{p}_i}, & i = 1, 2, \cdots, k, \\[2ex] N^{-1} + \dfrac{m}{nN}\dfrac{\hat{p}_{i-k}}{\hat{q}_{i-k}}, & i = k+1, \cdots, 2k. \end{cases}$$

PROOF. Our maximum likelihood estimation problem is the same as the one described by (5.5), (5.6) and (5.7) in Barlow and Brunk (1972) and they have shown that the solution also satisfies

$$\textstyle\sum_{j=1}^{i} p_j \ge \sum_{j=1}^{i} (m\hat{p}_j + n\hat{q}_j)/N \ge \sum_{j=1}^{i} q_j$$

for $i = 1, 2, \cdots, k-1$ with equality for $i = k$. Letting $t = (p_1/m\hat{p}_1, p_2/m\hat{p}_2, \cdots, p_k/m\hat{p}_k, q_1/n\hat{q}_1, \cdots, q_k/n\hat{q}_k)$ these restrictions are equivalent to

$$\textstyle\sum_{j=1}^{i} w_j(t_j - h_j) \ge 0 \text{ and } \sum_{j=k+1}^{k+i} w_j(h_j - t_j) \ge 0, \qquad i=1, 2, \cdots, k-1 \tag{2.4}$$

and

$$\textstyle\sum_{j=1}^{k} w_j(t_j - h_j) = 0 = \sum_{j=k+1}^{2k} w_j(h_j - t_j).$$

From Barlow and Brunk (1972), (2.4) is equivalent to $h - t \in B^{w*}$. Hence, with ϕ as before

$$\bar{t} = (\bar{p}_1/m\hat{p}_1, \bar{p}_2/m\hat{p}_2, \cdots, \bar{p}_k/m\hat{p}_k, \bar{q}_1/n\hat{q}_1, \cdots, \bar{q}_k/n\hat{q}_k)$$

solves:

$$\text{minimize } \textstyle\sum_{i=1}^{2k} w_i \phi(t_i) \quad \text{subject to} \quad h - t \in B^{w*}.$$

Appealing to Theorem 3.4 of Barlow and Brunk (1972) again, we have that $(\bar{p}, \bar{q}) = wE_w(h \mid B)$, which is the desired conclusion.

Since membership in B imposes no restrictions between the first k coordinates and the last k coordinates of a point, $(E(\cdot \mid B)_1, E(\cdot \mid B)_2, \cdots, E(\cdot \mid B)_k)$ and $(E(\cdot \mid B)_{k+1}, \cdots, E(\cdot \mid B)_{2k})$ can be computed independently. It follows that

$$\bar{p} = \hat{p} E_{\hat{p}} \left(\frac{m\hat{p} + n\hat{q}}{N\hat{p}} \,\middle|\, C \right) \tag{2.5}$$

and

$$\bar{q} = \hat{q} E_{\hat{q}} \left(\frac{m\hat{p} + n\hat{q}}{N\hat{q}} \,\middle|\, C' \right) = -\hat{q} E_{\hat{q}} \left(-\frac{m\hat{p} + n\hat{q}}{N\hat{q}} \,\middle|\, C \right),$$

where C' denote the cone $\{x: x_1 \le x_2 \le \cdots \le x_k\}$.

THEOREM 2.4. *If* $p \gg q$, *then* $P\{\lim_{m,n\to\infty} (\bar{p}, \bar{q}) = (p, q)\} = 1$.

PROOF. Since $E_w(g + e_k \mid C) = E_w(g \mid C) + e_k$, it follows from (2.5) that

$$\bar{p} - \hat{p} = (n/N)\hat{p} E_{\hat{p}} \left(\frac{\hat{q} - \hat{p}}{\hat{p}} \,\middle|\, C \right) \quad \text{and} \quad \bar{q} - \hat{q} = -(m/N)\hat{q} E_{\hat{q}} \left(\frac{\hat{q} - \hat{p}}{\hat{q}} \,\middle|\, C \right). \tag{2.6}$$

By the strong law of large numbers $P\{\lim_{m,n\to\infty} (\hat{p}, \hat{q}) = (p, q)\} = 1$. Since $(n/N)\hat{p}$ and $(m/N)\hat{q}$ are bounded and $E_w(x \mid C)$ is continuous in x and w, we need only show that

$$E_p \left(\frac{q - p}{p} \,\middle|\, C \right) = E_q \left(\frac{q - p}{q} \,\middle|\, C \right) = 0,$$

or equivalently

$$E_p \left(\frac{q}{p} \,\middle|\, C \right) = -E_q \left(-\frac{p}{q} \,\middle|\, C \right) = e_k.$$

In the proof of Theorem 2.2, it was shown that $p \gg q$ implies that $E_p(q/p \mid C) = e_k$ and the proof of $E_q(-p/q \mid C) = -e_k$ is similar.

It is interesting to observe that $(\bar{p}, \bar{q})$ is strongly consistent for (p, q) for any sequence of sample sizes (m, n) provided m and n simultaneously approach ∞.

3. Tests with a known standard: one-sample tests. In this and the next section we use λ generically to denote the likelihood ratio. Suppose q is known and that we have a random sample of size m from the p-population and consider testing $H_0: p = q$ against $H_1 - H_0$ where $H_1: p \gg q$. Let

$$S_{01} = -2 \ln \lambda = -2m \textstyle\sum_{i=1}^k \bar{p}_i (\ln q_i - \ln \bar{p}_i).$$

Since H_0 is a boundary point of H_1 the usual limiting chi-squared results for $-2 \ln \lambda$ do not apply. However the next result shows that the limit distribution is a mixture of chi-squared distributions. Before stating the result we define the mixing proportions. Let $w = (w_1, w_2, \cdots, w_k)$ be positive weights and let $W_1, W_2, \cdots, W_k$ be independent normal variables with zero means and variances $w_1^{-1}, w_2^{-1}, \cdots, w_k^{-1}$ respectively. We denote the probability that $E_w(W \mid C)$ has exactly ℓ distinct values (level sets) by $P_w(\ell, k)$.

THEOREM 3.1. *If* H_0 *is true then for any real number* t

$$\lim_{m\to\infty} P(S_{01} \ge t) = \textstyle\sum_{\ell=1}^k P_q(\ell, k) P(\chi^2_{k-\ell} \ge t),$$

where χ^2_ν *is a chi-squared variable with* ν *degrees freedom* $(\chi^2_0 \equiv 0)$.

PROOF. Writing a second order Taylor's expansion for $\ln q_i$ and $\ln \bar{p}_i$ about the point

$\hat{p}_i$, S_{01} can be expressed as follows:

(3.1) $$S_{01} = \sum_{i=1}^k \hat{p}_i\alpha_i^{-2}\{\sqrt{m}(\hat{p}_i - q_i)\}^2 - \sum_{i=1}^k \hat{p}_i\beta_i^{-2}\{\sqrt{m}(\bar{p}_i - \hat{p}_i)\}^2,$$

where α_i is between $\hat{p}_i$ and q_i and β_i is between $\bar{p}_i$ and $\hat{p}_i$. Let $U_1, U_2, \cdots, U_k$ be independent normal variables which are centered at their expectations and have variances $p_1^{-1}, p_2^{-1}, \cdots, p_k^{-1}$, respectively. Then the random vector $\sqrt{m}(\hat{p} - p)$ converges in distribution to $(p_1(U_1 - \tilde{U}), p_2(U_2 - \tilde{U}), \cdots, p_k(U_k - \tilde{U}))$ where $\tilde{U} = \sum_{i=1}^k p_i U_i$ (cf. Robertson, 1978). Hence, appealing to Theorem 4.4 of Billingsley (1968), we have that

$$(\sqrt{m}(\hat{p} - p), \hat{p}, \bar{p}, \alpha, \beta) \to_{\mathscr{D}} (p_1(U_1 - \tilde{U}), \cdots, p_k(U_k - \tilde{U}), p, p, p, p)$$

provided H_0 is true. Thus, under H_0, S_{01} converges in distribution to

(3.2) $$\sum_{i=1}^k q_i(U_i - \tilde{U})^2 - \sum_{i=1}^k q_i\{E_q(\tilde{U}e_k - U \mid C)_i\}^2.$$

Now, noting that $E_q(\tilde{U}e_k - U \mid C) = \tilde{U}e_k + E_q(-U \mid C)$, squaring the binomials in (3.2), combining terms and using Theorem 7.8 of Barlow *et al.* (1972), (3.2) can be rewritten as

$$\sum_{i=1}^k q_i\{E_q(W \mid C)_i - W_i\}^2$$

where $W_i = -U_i$; $i = 1, 2, \cdots, k$. Corollary 2.6 of Robertson and Wegman (1978) gives the desired conclusion.

If $q_1 = q_2 = \cdots = q_k = k^{-1}$ then the $P(\ell, k)$ can be determined recursively from Corollary B on page 145 of Barlow *et al.* (1972). Their Appendix A5 gives the $P(\ell, k)$ for $k \le 12$ in this case. However, if the q_i are not all equal the $P(\ell, k)$ are much more difficult to compute. Equation (3.23) of Barlow *et al.* (1972) is a recursive relation from which one can obtain the $P(\ell, k)$ provided $P(j, j)$ is known for $j \le k$. Barlow *et al.* (1972) contains closed form expressions for $P(j, j)$ with $j \le 4$ and the tables in Abrahamson (1964) can be used to compute $P(5, 5)$. Robertson and Wright (1980) have obtained bounds for certain chi-bar-squared distributions. Their results show that

(3.3) $$\lim_{m\to\infty} P(S_{01} \ge t) \le \{P(\chi_{k-1}^2 \ge t) + P(\chi_{k-2}^2 \ge t)\}/2$$

and of course, one could obtain a conservative test using the upper bound in (3.3).

Next, we consider the one-sample likelihood ratio test of H_1 versus H_2. The test statistic is

$$S_{12} = -2 \ln \lambda = -2m \sum_{i=1}^k \hat{p}_i(\ln \bar{p}_i - \ln \hat{p}_i).$$

Let $P_p(E)$ denote the probability of the event E computed under the assumption that p is the population vector of probabilities.

THEOREM 3.2. *For any p satisfying H_1,* i.e. $p \gg q$, *and for all t*

$$\lim_{m\to\infty} P_p(S_{12} \ge t) \le \lim_{m\to\infty} P_q(S_{12} \ge t)$$

and

$$\lim_{m\to\infty} P_q(S_{12} \ge t) = \sum_{\ell=1}^k P_q(\ell, k) P(\chi_{\ell-1}^2 \ge t).$$

PROOF. Writing a second order Taylor's expansion for $\ln \bar{p}_i$ about the point $\hat{p}_i$ we see that S_{12} can be written

(3.4) $$\sum_{i=1}^k \hat{p}_i\gamma_i^{-2}\{\sqrt{m}(\bar{p}_i - \hat{p}_i)\}^2,$$

where γ_i is between $\bar{p}_i$ and $\hat{p}_i$. Now we want to obtain the limiting distribution of (3.4) under H_1 and to show that this limit is stochastically largest for $p = q$. Let $p \gg q$ and let $0 = \eta_0 < \eta_1 < \cdots < \eta_A = k$ with $p_1 + \cdots + p_i = q_1 + \cdots + q_i$ for $i = \eta_1, \eta_2, \cdots, \eta_A$ and $p_1 + \cdots + p_i > q_1 + \cdots + q_i$ for $i \ne \eta_1, \eta_2, \cdots, \eta_A$. By the strong law of large numbers, for almost all ω (in the underlying probability space) there is an $m_0(\omega)$ and an $\varepsilon > 0$ for which

$$(q_{\eta_j+1} + \cdots + q_i)/(\hat{p}_{\eta_j+1} + \cdots + \hat{p}_i) < 1 - \varepsilon$$

for each $j = 0, \cdots, A-1$ and $i > \eta_j$ with $i \neq \eta_{j+1}, \cdots, \eta_A$ and

$$(q_{\eta_j+1} + \cdots + q_{\eta_\ell})/(\hat{p}_{\eta_{j+1}} + \cdots + \hat{p}_{\eta_\ell}) > 1 - \varepsilon$$

for each $0 \leq j < \ell \leq A$ provided $m \geq m_0(\omega)$. So in using the LSA to compute $E_{\hat{p}}(q/\hat{p} - e_k \mid C)$ for such an ω and m, we see that the level sets are of the form $\{\eta_j + 1, \cdots, \eta_\ell\}$ with $0 \leq j < \ell \leq A$. Consider the closed, convex cone

$$D = \{v \in C: v_1 = \cdots = v_{\eta_1}, v_{\eta_1+1} = \cdots = v_{\eta_2}, \cdots, v_{\eta_{A-1}+1} = \cdots = v_{\eta_A}\}.$$

If $E_w(g \mid D)$ denotes the projection of g onto D with respect to the distance associated with $(\cdot, \cdot)_w$, then for such ω and m

$$E_{\hat{p}}(q/\hat{p} - e_k \mid C) = E_{\hat{p}}(q/\hat{p} - e_k \mid D) \tag{3.5}$$

since $E_{\hat{p}}(q/\hat{p} - e_k \mid C) \in D$. One way to compute $E_w(g \mid D)$ is to first obtain g^* which is constant on $\{\eta_j + 1, \cdots, \eta_{j+1}\}$ by setting

$$g_i^* = \textstyle\sum_{\ell=\eta_j+1}^{\eta_{j+1}} w_\ell g_\ell / \sum_{\ell=\eta_j+1}^{\eta_{j+1}} w_\ell$$

for $i = \eta_j + 1, \cdots, \eta_{j+1}$ and $j = 0, 1, \cdots, A-1$, and then to apply the LSA to g^* with weights $w_1, \cdots, w_k$. (For any vector $g \in R^k$, the notation g^* will be used only to denote the vector defined in this manner.) If $g = q/\hat{p} - e_k$, $f = p/\hat{p} - e_k$ and $w = \hat{p}$, then $g^* = f^*$ and hence $E_{\hat{p}}(q/\hat{p} - e_k \mid D) = E_{\hat{p}}\{(p - \hat{p})/\hat{p} \mid D\}$. Clearly, $(\sqrt{m}\,(\hat{p} - p), \hat{p}, \gamma)$ converges in distribution to $(p_1(U_1 - \tilde{U}), p_2(U_2 - \tilde{U}), \cdots, p_k(U_k - \tilde{U}), p, p)$ with $U_1, \cdots, U_k$ and $\tilde{U}$ defined as before. Using (2.1), we see that (3.4) converges in distribution to

$$\textstyle\sum_{i=1}^k p_i\{E_p(\tilde{U}e_k - U \mid D)_i\}^2 = \sum_{i=1}^k p_i\{E_p(W \mid D)_i - \tilde{W}\}^2 \tag{3.6}$$

where $W_i = -U_i$ for $i = 1, 2, \cdots, k$ and $\tilde{W} = \sum_{i=1}^k p_i W_i$. Since $E_p(W \mid D)$ is constant on $\{\eta_j + 1, \cdots, \eta_{j+1}\}$ for $j = 0, 1, \cdots, A-1$, (3.6) can be written as $\sum_{i=1}^k q_i\{E_p(W \mid D)_i - \tilde{W}\}^2$. To compute $E_p(W \mid D)$ we first obtain

$$W_i^* = \textstyle\sum_{\ell=\eta_j+1}^{\eta_{j+1}} p_\ell W_\ell / \sum_{\ell=\eta_j+1}^{\eta_{j+1}} p_\ell$$

for $i = \eta_j + 1, \cdots, \eta_{j+1}$ and $j = 0, 1, \cdots, A-1$. Now, if

$$T_i = (p_i/q_i) W_i, \; \tilde{T} = \textstyle\sum_{i=1}^k q_i T_i \quad \text{and} \quad T_i^* = \sum_{\ell=\eta_j+1}^{\eta_{j+1}} q_\ell T_\ell / \sum_{\ell=\eta_j+1}^{\eta_{j+1}} q_\ell,$$

then $W^* = T^*$. Since $\tilde{W} = \sum_{i=1}^k q_i W_i^* = \tilde{T}$ and $E_p(W \mid D) = E_p(W^* \mid D) = E_q(W^* \mid D) = E_q(T \mid D)$, (3.6) is equal to $\sum_{i=1}^k q_i\{E_q(T \mid D)_i - \tilde{T}\}^2$. However,

$$\textstyle\sum_{i=1}^k q_i(T_i - \tilde{T})^2 = \sum_{i=1}^k q_i\{T_i - E_q(T \mid D)_i\}^2 + \sum_{i=1}^k q_i\{E_q(T \mid D)_i - \tilde{T}\}^2.$$

The first term on the right hand side of the previous equation is $\| T - E_q(T \mid D) \|_q^2$ which is smallest when D is largest, that is $D = C$, which occurs if $p = q$. So the first conclusion of the theorem is established and it follows that $p = q$ is the asymptotically least favorable distribution in H_1, in the sense that the probability of a type I error for the asymptotic test is largest if $p = q$. Furthermore, if $p = q$, then $D = C$ and the second conclusion is a consequence of a result due to Bartholomew (cf. Theorem 3.1 of Barlow *et al.*, 1972). The proof is completed.

As we have noted earlier, the computation of the $P(\ell, k)$ may be tedious and so we apply the bounds for chi-bar-squared distributions given in Robertson and Wright (1980) to obtain

$$\sup_{p \in H_1} \lim_{m \to \infty} P_p(S_{12} \geq t) \leq \textstyle\sum_{\ell=1}^k \binom{k-1}{\ell-1} 2^{-k+1} P(\chi_{\ell-1}^2 \geq t). \tag{3.7}$$

Of course, the upper bound in (3.7) can be used to determine a conservative asymptotic test.

If q is known one might want to test $H_0: p = q$ versus $H_1': q \gg p$. However, if we define

$$p' = (p_k, p_{k-1}, \cdots, p_1) \quad \text{and} \quad q' = (q_k, q_{k-1}, \cdots, q_1)$$

then $p' \gg q'$ is equivalent to $q \gg p$. So the tests developed in this section can be used to test $H_0: p = q$ versus $H_1': q \gg p$ and H_1' versus H_2 with q known.

4. Two-sample tests. In this section we suppose that $\hat{p}$ and $\hat{q}$ are the relative frequencies of successes corresponding to independent random samples of size m and n from the p and q populations respectively. We first consider the likelihood ratio test of H_0 versus $H_1 - H_0$ where $H_1: p \gg q$. The test statistic, $-2 \ln \lambda$, can be expressed as

$$T_{01} = 2m \textstyle\sum_{i=1}^k \hat{p}_i(\ln \bar{p}_i - \ln p_i^0) + 2n \sum_{i=1}^k \hat{q}_i(\ln \bar{q}_i - \ln q_i^0) \tag{4.1}$$

where $p_i^0 = q_i^0 = (m\hat{p}_i + n\hat{q}_i)/N$; $i = 1, 2, \cdots, k$ and $\bar{p}$ and $\bar{q}$ are given by Theorem 2.3.

THEOREM 4.1. *If $p = q$, then for each real t*

$$\lim_{m,n\to\infty} P(T_{01} \ge t) = \textstyle\sum_{\ell=1}^k P_p(\ell, k) P(\chi_{k-\ell}^2 \ge t).$$

Furthermore,

$$\sup_{p=q} \lim_{m,n\to\infty} P(T_{01} \ge t) = \tfrac{1}{2}\{P(\chi_{k-1}^2 \ge t) + P(\chi_{k-2}^2 \ge t)\}.$$

PROOF. Writing a second order Taylor's expansion for $\ln \bar{p}_i$ and $\ln p_i^0$ about $\hat{p}_i$ and for $\ln \bar{q}_i$ and $\ln q_i^0$ about $\hat{q}_i$, we see that T_{01} can be written as the sum of

$$\textstyle\sum_{i=1}^k \hat{p}_i \theta_i^{-2}\{\sqrt{m}(p_i^0 - \hat{p}_i)\}^2 - \sum_{i=1}^k \hat{p}_i \nu_i^{-2}\{\sqrt{m}(\bar{p}_i - \hat{p}_i)\}^2 \tag{4.2}$$

and

$$\textstyle\sum_{i=1}^k \hat{q}_i \rho_i^{-2}\{\sqrt{n}(q_i^0 - \hat{q}_i)\}^2 - \sum_{i=1}^k \hat{q}_i \sigma_i^{-2}\{\sqrt{n}(\bar{q}_i - \hat{q}_i)\}^2, \tag{4.3}$$

where $\theta_i(\nu_i)$ is between p_i^0 and $\hat{p}_i(\bar{p}_i$ and $\hat{p}_i)$ and $\rho_i(\sigma_i)$ is between q_i^0 and $\hat{q}_i(\bar{q}_i$ and $\hat{q}_i)$. Let $V = (V_1, V_2, \cdots, V_k)$ with the V_i independent normal variables which have zero means and variances $q_1^{-1}, q_2^{-1}, \cdots, q_k^{-1}$, respectively, and suppose that V is independent of the U defined in the previous section. If we set $\tilde{V} = \sum_{i=1}^k q_i V_i$, then as m and n simultaneously approach ∞

$$(\sqrt{m}(\hat{p} - p), \sqrt{n}(\hat{q} - q)) \\ \to_{\mathscr{D}} (p_1(U_1 - \tilde{U}), \cdots, p_k(U_k - \tilde{U}), q_1(V_1 - \tilde{V}), \cdots, q_k(V_k - \tilde{V})).$$

Furthermore, since $\bar{p}_i$ and $\hat{p}_i(\bar{q}_i$ and $\hat{q}_i)$ are strongly consistent for $p_i(q_i)$ provided $p \gg q$, it follows that, with probability one, $\theta = (\theta_1, \cdots, \theta_k) \to p$, $\nu = (\nu_1, \cdots, \nu_k) \to p$, $\rho = (\rho_1, \cdots, \rho_k) \to q$ and $\sigma = (\sigma_1, \cdots, \sigma_k) \to q$.

Let $p = q$ and $m, n \to \infty$ so that $m/N \to a \in [0, 1]$. Since (4.2) and (4.3) are continuous functions of $(\sqrt{m}(\hat{p} - p), \sqrt{n}(\hat{q} - q), \hat{p}, \hat{q}, \theta, \nu, \rho, \sigma)$, we may apply the weak convergence results mentioned earlier to show that (4.2)′ converges in distribution to the product of $(1 - a)$ and

$$\textstyle\sum_{i=1}^k p_i[\{\sqrt{a}(V_i - \tilde{V}) - \sqrt{1-a}(U_i - \tilde{U})\}^2 \\ - \{E_p(\sqrt{a}(V - \tilde{V}e_k) - \sqrt{1-a}(U - \tilde{U}e_k) \mid C)_i\}^2] \tag{4.4}$$

and (4.3) converges in distribution to the product of a and (4.4). Hence, T_{01} converges to (4.4), which can be written as

$$\textstyle\sum_{i=1}^k p_i[\{(\sqrt{a}V_i - \sqrt{1-a}U_i) + (\sqrt{1-a}\tilde{U} - \sqrt{a}\tilde{V})\}^2 \\ - \{E_p(\sqrt{a}V - \sqrt{1-a}U \mid C)_i + (\sqrt{1-a}\tilde{U} - \sqrt{a}\tilde{V})\}^2].$$

Squaring the binomials in the above expression and applying Theorem 7.8 of Barlow *et al.* (1972), this expression can be written as

(4.5) $$\sum_{i=1}^{k} p_i[W_i^2 - \{E_p(W \mid C)_i\}^2] = \sum_{i=1}^{k} p_i\{W_i - E_p(W \mid C)_i\}^2,$$

where $W_i = \sqrt{a}V_i - \sqrt{1-a}\,U_i \sim N(0, p_i^{-1})$ and $W_1, W_2, \cdots, W_k$ are independent. Since the limit (expression (4.5)) does not depend on a, T_{01} converges in distribution to (4.5) for any sequence of m and n's which both approach infinity (cf. Theorem 2.3 of Billingsley, 1968). As we have seen earlier (4.5) has the chi-bar-squared distribution stated in the first conclusion of the theorem. The second conclusion follows from the results given in Robertson and Wright (1980a).

In this two-sample situation the vector p is not specified by H_0. One could use $p^0 = (p_1^0, \cdots, p_k^0)$ as an estimate of the unknown p and compute the $P(\ell, k)$ based on this estimate. The use of the resulting chi-bar-squared distribution would provide an approximate large sample test. Or, if one wanted an asymptotic test with size α, the test could be based on the second conclusion of Theorem 4.1, that is the critical value, C, could be chosen to satisfy $P(\chi_{k-1}^2 \geq C) + P(\chi_{k-2}^2 \geq C) = 2\alpha$.

Next, consider a likelihood ratio test of H_1 versus $H_2 = \sim H_1$. The test statistic, $T_{12} = -2 \ln \lambda$, can be written as

(4.6) $$T_{12} = -2m \sum_{i=1}^{k} \hat{p}_i(\ln \bar{p}_i - \ln \hat{p}_i) - 2n \sum_{i=1}^{k} \hat{q}_i(\ln \bar{q}_i - \ln \hat{q}_i).$$

THEOREM 4.2. *If $P_{p,q}(E)$ denotes the probability of the event E computed under the assumption that p and q are the values of the parameters, then for each real t*

$$\sup_{p \gg q} \lim_{m,n\to\infty} P_{p,q}(T_{12} \geq t) = \sup_{p=q} \lim_{m,n\to\infty} P_{p,q}(T_{12} \geq t)$$

and

$$\sup_{p=q} \lim_{m,n\to\infty} P(T_{12} \geq t) = \sum_{\ell=1}^{k} \binom{k-1}{\ell-1} 2^{-k+1} P(\chi_{\ell-1}^2 \geq t).$$

PROOF. Writing a second order expansion for $\ln \bar{p}_i$ ($\ln \bar{q}_i$) about the point $\hat{p}_i(\hat{q}_i)$ and expressing $\bar{p}_i - \hat{p}_i$ and $\bar{q}_i - \hat{q}_i$ as projections, (4.6) becomes

(4.7) $$(n/N) \sum_{i=1}^{k} \hat{p}_i^3 \tau_i^{-2} \left\{ \sqrt{mn/N}\, E_{\hat{p}}\left(\frac{\hat{q}-\hat{p}}{\hat{p}} \,\middle|\, C \right)_i \right\}^2 + (m/N) \sum_{i=1}^{k} \hat{q}_i^3 \phi_i^{-2} \left\{ \sqrt{mn/N}\, E_{\hat{q}}\left(\frac{\hat{q}-\hat{p}}{\hat{q}} \,\middle|\, C \right)_i \right\}^2,$$

where $\tau_i(\phi_i)$ is between $\bar{p}_i$ and $\hat{p}_i$ ($\bar{q}_i$ and $\hat{q}_i$). Define D as in the one-sample situation. As in that case, for almost all ω and m and n sufficiently large

$$E_{\hat{p}}\left(\frac{\hat{q}-\hat{p}}{\hat{p}} \,\middle|\, C \right) = E_{\hat{p}}\left(\frac{\hat{q}-\hat{p}}{\hat{p}} \,\middle|\, D \right) \text{ and } E_{\hat{q}}\left(\frac{\hat{q}-\hat{p}}{\hat{q}} \,\middle|\, C \right) = E_{\hat{q}}\left(\frac{\hat{q}-\hat{p}}{\hat{q}} \,\middle|\, D \right).$$

Considering the algorithm for computing $E_w(g \mid D)$, we see that

$$E_{\hat{p}}\left(\frac{\hat{q}-\hat{p}}{\hat{p}} \,\middle|\, D \right) = E_{\hat{p}}\left\{ \frac{(\hat{q}-q)-(\hat{p}-p)}{\hat{p}} \,\middle|\, D \right\} \text{ and } E_{\hat{q}}\left(\frac{\hat{q}-\hat{p}}{\hat{q}} \,\middle|\, D \right) = E_{\hat{q}}\left\{ \frac{(\hat{q}-q)-(\hat{p}-p)}{\hat{q}} \,\middle|\, D \right\}.$$

Hence, if $m, n \to \infty$ with $m/N \to a \in [0, 1]$, then, with U, V, $\tilde{U}$ and $\tilde{V}$ defined as before, (4.7) converges to

(4.8) $$\sum_{i=1}^{k} q_i \left[E_q\left\{ \frac{q\sqrt{a}(V - \tilde{V}e_k) - p\sqrt{1-a}(U - \tilde{U}e_k)}{q} \,\middle|\, D \right\}_i \right]^2$$

since $E_p(f/p \mid D) = E_q(f/q \mid D)$ for any f defined on $\{1, 2, \cdots, k\}$. If $T_i = (p_i/q_i)U_i$ and $\tilde{T} = \sum_{i=1}^{k} q_i T_i$ then $T^* = U^*$, $\tilde{T} = \tilde{U}$ and (4.8) can be written as

(4.9) $$\sum_{i=1}^{k} q_i\{E_q(W \mid D)_i - \tilde{W}\}^2$$

with

$$W_i = \sqrt{a}\, V_i - \sqrt{1-a}\, T_i \quad \text{and} \quad \tilde{W} = \textstyle\sum_{i=1}^k q_i W_i.$$

Since (4.9) does not depend on a, T_{12} converges in distribution to (4.9) for any sequence of m and n's which simultaneously approach ∞. As before, (4.9) is made stochastically largest by setting $D = C$ or $p = q$. So the first conclusion of the theorem is established. In this case, that is $p = q$ or $D = C$, (4.9) has a chi-bar-squared distribution with tail probabilities $\sum_{\ell=1}^k P_p(\ell, k) P(\chi^2_{\ell-1} \geq t)$ for all real t. The second conclusion of the theorem follows from Theorem 1 and Remark 2 of Robertson and Wright (1980a).

5. Summary. We outline below the procedures that have been developed here for testing $H_0: p = q$ vs. $H_1 - H_0$ where $H_1: p \gg q$ and H_1 vs. $H_2 = \sim H_1$.

I. *One-sample tests*: *known standard.* $\hat{p}$ is the relative frequency estimate of p based on a sample of size m and q is known.

A.
M.l.e. of p subject to $p \gg q: \bar{p} = \hat{p} E_{\hat{p}}(q/\hat{p} \mid C)$ where $E_{\hat{p}}(\cdot \mid C)$ can be computed by the LSA or

$$\bar{p}_i = \hat{p}_i \min_{1 \leq \alpha \leq i} \max_{i \leq \beta \leq k} \textstyle\sum_{j=\alpha}^{\beta} q_i / \sum_{j=\alpha}^{\beta} \hat{p}_j, \; i = 1, 2, \cdots, k.$$

B. Test of H_0 vs. $H_1 - H_0$.

(1) Test statistic:

$$S_{01} = -2 \ln \lambda = 2m \textstyle\sum_{i=1}^k \hat{p}_i (\ln \bar{p}_i - \ln q_i)$$

(2) Null distribution:

$$\lim_{m\to\infty} P(S_{01} \geq t) = \textstyle\sum_{\ell=1}^k P_q(\ell, k) P(\chi^2_{k-\ell} \geq t).$$

C. Test of H_1 vs. H_2

(1) Test statistic:

$$S_{12} = -2 \ln \lambda = 2m \textstyle\sum_{i=1}^k \hat{p}_i (\ln \hat{p}_i - \ln \bar{p}_i)$$

(2) Null distribution:

$$\sup_{p \gg q} \lim_{m\to\infty} P(S_{12} \geq t) = \textstyle\sum_{\ell=1}^k P_q(\ell, k) P(\chi^2_{\ell-1} \geq t).$$

II. *Two-sample tests.* $\hat{p}$ and $\hat{q}$ are independent relative frequency estimates of p and q based on samples of size m and n respectively.

A. M.l.e. of p and q subject to $p \gg q: \bar{p} = (\hat{p}/N) E_{\hat{p}} \{(m\hat{p} + n\hat{q})/\hat{p} \mid C\}$ and $\bar{q} = -(\hat{q}/N) E_{\hat{q}} \{-(m\hat{p} + n\hat{q})/\hat{q} \mid C\}$ with the projections computed by the LSA or

$$\bar{p}_i = (\hat{p}_i/N) \min_{1 \leq \alpha \leq i} \max_{i \leq \beta \leq k} \textstyle\sum_{j=\alpha}^{\beta} (m\hat{p}_j + n\hat{q}_j) / \sum_{j=\alpha}^{\beta} \hat{p}_i.$$

$$\hat{q}_i = (\hat{q}_i/N) \max_{1 \leq \alpha \leq i} \min_{i \leq \beta \leq k} \textstyle\sum_{j=\alpha}^{\beta} (m\hat{p}_j + n\hat{q}_j) / \sum_{j=\alpha}^{\beta} \hat{q}_j.$$

B. Test of H_0 vs. $H_1 - H_0$.

(1) Test statistic:

$$T_{01} = -2 \ln \lambda = 2m \textstyle\sum_{i=1}^k \hat{p}_i (\ln \bar{p}_i - \ln p_i^0) + 2n \sum_{i=1}^k \hat{q}_i (\ln \bar{q}_i - \ln q_i^0)$$

with $p_1^0 = q_1^0 = (m\hat{p}_i + n\hat{q}_i)/N$.

(2) Null distribution:

$$\sup_{p=q} \lim_{m,n\to\infty} P(T_{01} \geq t) = \tfrac{1}{2}\{P(\chi^2_{k-1} \geq t) + P(\chi^2_{k-2} \geq t)\}.$$

C. Test of H_1 vs. H_2.

(1) Test statistic:

$$T_{12} = -2 \ln \lambda = 2m \sum_{i=1}^{k} \hat{p}_i(\ln \hat{p}_i - \ln \bar{p}_i) + 2n \sum_{i=1}^{k} \hat{q}_i(\ln \hat{q}_i - \ln \bar{q}_i).$$

(2) Null distribution:

$$\sup_{p \gg q} \lim_{m,n\to\infty} P(T_{12} \geq t) = \sum_{\ell=1}^{k} \binom{k-1}{\ell-1} 2^{-k+1} P(\chi^2_{\ell-1} \geq t).$$

Acknowledgements. We thank the referees and the editors for their efforts. One referee's observation resulted in an improvement of the proofs of Theorems 3.2 and 4.2.

REFERENCES

ABRAHAMSON, I. G. (1964). Orthant probabilities for the quadrivariate normal distribution. *Ann. Math. Statist.* **35** 1685–1703.

BARLOW, R. E., BARTHOLOMEW, D. J., BREMNER, J. M. and BRUNK, H. D. (1972). *Statistical Inference Under Order Restrictions.* Wiley, New York.

BARLOW, R. E. and BRUNK, H. D. (1972). The isotonic regression problem and its dual. *J. Amer. Statist. Assoc.* **67** 140–147.

BILLINGSLEY, P. (1968). *Convergence of Probability Measures.* Wiley, New York.

BRUNK, H. D. (1965). Conditional expectation given a σ-lattice and applications. *Ann. Math. Statist.* **36** 1339–1350.

CHACKO, V. J. (1966). Modified chi-square test for ordered alternatives. *Sankhyā Ser. B* **28** 185–190.

GROVE, D. M. (1980). A test of independence against a class of ordered alternatives in a 2 × C contingency table. *J. Amer. Statist. Assoc.* **75** 454–459.

ROBERTSON, T. (1978). Testing for and against an order restriction on multinomial parameters. *J. Amer. Statist. Assoc.* **73** 197–202.

ROBERTSON, T. and WEGMAN, E. J. (1978). Likelihood ratio tests for order restrictions in exponential families. *Ann. Statist.* **6** 485–505.

ROBERTSON, T. and WRIGHT, F. T. (1980a). Bounds on mixtures of distributions arising in order restricted inference. Technical Report 69, Univ. of Iowa.

ROBERTSON, T. and WRIGHT F. T. (1980b). On measuring the conformity of a parameter set to a trend, with applications. Technical Report 71, Univ. of Iowa.

DEPARTMENT OF STATISTICS
THE UNIVERSITY OF IOWA
IOWA CITY, IOWA 52242

326 MATHEMATICS-COMPUTER SCIENCE BUILDING
UNIVERSITY OF MISSOURI, ROLLA
ROLLA, MISSOURI 65401