The Annals of Statistics
1978, Vol. 6, No. 4, 904–909

TESTING A LINEAR CONSTRAINT FOR MULTINOMIAL CELL FREQUENCIES AND DISEASE SCREENING[1]

BY JEROME H. KLOTZ

University of Wisconsin at Madison

For comparing two disease screening procedures with economic costs assigned to administration, false positives, and false negatives, the problem of testing a linear cell frequency constraint $\sum_{i=1}^{K} a_i p_i \leqq 0$ arises with the multinomial $(n, (p_1, p_2, \cdots, p_K))$ model. An ad hoc statistic based upon the estimate of the p_i values, $\sum_{i=1}^{K} a_i X_i/n$, is compared with the likelihood ratio statistic $-2 \ln \lambda$, the latter having an interesting form. For local (contiguous) alternatives the two statistics have similar large sample properties. However, the likelihood ratio statistic has greater large deviation efficiency for fixed alternatives and is recommended.

1. Introduction and notation. We observe a multinomial $(n, \mathbf{p})$, $\mathbf{p} = (p_1, p_2, \cdots, p_K)$ random vector $\mathbf{X} = (X_1, X_2, \cdots, X_K)$,

$$P[\mathbf{X} = \mathbf{x}] = n! \prod_{i=1}^{K} p_i^{x_i}/x_i! \,, \tag{1.1}$$

and consider the hypothesis

$$H_0: \sum_{i=1}^{K} a_i p_i \leqq 0 \quad \text{against} \quad H_1: \sum_{i=1}^{K} a_i p_i > 0 \tag{1.2}$$

where a_i are known constants $a_1 \leqq a_2 \leqq \cdots \leqq a_K$ and $a_1 < 0 < a_K$ to avoid trivialities.

Following Hoeffding (1965) denote $\Omega = \{p \mid p_i \geqq 0, \sum_i p_i = 1\}$, $\Omega_0 = \{\mathbf{p} \mid p_i > 0, \sum_i p_i = 1\}$, and the Kullback–Leibler information quantities by:

$$\begin{aligned} I(\boldsymbol{\nu}, \mathbf{p}) &= \sum_{i=1}^{K} \nu_i \ln (\nu_i/p_i) \\ I(A, \mathbf{p}) &= \inf \{I(\boldsymbol{\nu}, \mathbf{p}) \mid \boldsymbol{\nu} \in A\} \\ I(\boldsymbol{\nu}, \Lambda) &= \inf \{I(\boldsymbol{\nu}, \mathbf{p}) \mid \mathbf{p} \in \Lambda\} \\ I(A, \Lambda) &= \inf \{I(\boldsymbol{\nu}, \mathbf{p}) \mid \boldsymbol{\nu} \in A, \mathbf{p} \in \Lambda\} \end{aligned} \tag{1.3}$$

using the convention $0 \ln 0 = 0$.

2. An ad hoc test of *H*. Replacing the p_i in (1.2) by the estimators X_i/n, consider the statistic $\sum_{i=1}^{n} a_i X_i/n$ which has variance

$$(\textstyle\sum_i a_i^2 p_i q_i/n) - (2 \sum_{i<j} \sum a_i a_j p_i p_j/n) = (\sum_i a_i^2 p_i/n) - ([\sum_i a_i p_i]^2/n) \tag{2.1}$$

where $q_i = 1 - p_i$. On the boundary of H_0 and H_1, $\sum_i a_i p_i = 0$ so that (2.1) reduces to $\sum_i a_i^2 p_i/n$. Thus the statistic defined for nonzero denominator by

$$(n^{\frac{1}{2}} \textstyle\sum_i a_i X_i/n)(\sum_i a_i^2 X_i/n)^{-\frac{1}{2}} \tag{2.2}$$

Received March 1976; revised August 1977.

[1] Research supported by National Institute of Health Grant 1 R01 CA 18332-01.

AMS 1970 *subject classifications.* Primary 62F05; Secondary 62P10.

Key words and phrases. Testing, linear constraint, multinomial, large deviations, efficiency, information, likelihood ratio test, disease screening.

will have an asymptotic standard normal distribution on the boundary using consistency ($X_i/n \to_p p_i$) and the multivariate central limit theorem for the multinomial.

3. The likelihood ratio statistic. Denote $\Lambda = \{\mathbf{p} \mid \sum p_i = 1, p_i \geqq 0, \sum a_i p_i \leqq 0\}$ for H_0; then the likelihood ratio statistic $-2 \ln \lambda$ is given by $2nI(\mathbf{X}/n, \Lambda)$ using (1.1) and (1.3). The constrained minimization can be explicity solved using Kuhn–Tucker nonlinear programming methods (e.g., Mangasarian (1969), page 94). The solution is verified with the derivation avoided in the proof of the following:

THEOREM 1. *Define* $a_*(\mathbf{X}) = \min (a_i \mid x_i > 0)$, *then*

$$I(\mathbf{X}/n, \Lambda) = 0 \,, \qquad \text{if } \sum a_i X_i/n \leqq 0 \tag{3.1a}$$

$$= \sum_i \frac{X_i}{n} \ln (1 + \hat{\eta} a_i) \,, \qquad \text{if } \sum a_i \frac{X_i}{n} > 0 \,, \quad a_*(\mathbf{X}) < 0 \tag{3.1b}$$

$$= \sum_i \frac{X_i}{n} \ln ((a_1 - a_i)/a_1) \,, \qquad \text{if } \sum a_i \frac{X_i}{n} > 0 \,, \quad a_*(X) \geqq 0 \tag{3.1c}$$

where for case (3.1 b) $\hat{\eta}(\mathbf{X})$ *is the unique root of the equation*

$$\sum_i X_i/[n(1 + \hat{\eta} a_i)] = 1 \,, \quad 0 < \hat{\eta} < -1/a_*(\mathbf{X}) \,. \tag{3.2}$$

PROOF. For the case (3.1 a) $\hat{p}_i = x_i/n$ minimizes $I(\mathbf{X}/n, \mathbf{p})$. The convexity of $u \ln u$ yields

$$\sum \frac{X_i}{n} \ln \frac{X_i}{np_i} \geqq \left(\sum_i p_i \frac{X_i}{np_i}\right) \ln \left(\sum_j p_j \frac{X_j}{np_j}\right) = 0 \,,$$

using $\sum_i p_i = 1$, $\sum_i p_i X_i/(np_i) = 1$ and noting that the constraint satisfied by $\hat{p}_i$ gives the value zero.

For the case (3.1 b) let $\mathbf{p} \in \Lambda$ and denote $\theta_i = p_i(1 + \hat{\eta} a_i)$, and $u_i = X_i/[np_i(1 + \hat{\eta} a_i)]$ where $\hat{\eta}$ satisfies (3.2). Then

$$\sum_i \frac{X_i}{n} \ln \frac{X_i}{np_i} = \sum_i \frac{X_i}{n} \ln (1 + \hat{\eta} a_i) + \sum_i \theta_i u_i \ln u_i \,. \tag{3.3}$$

Using the convexity of $u \ln u$, setting $\zeta_i = \theta_i / \sum_{k=1}^{K} \theta_k$, gives

$$\sum_i \zeta_i u_i \ln u_i \geqq \left(\sum_j \zeta_j u_j\right) \ln \left(\sum_i \zeta_i u_i\right) \tag{3.4}$$

so that $\sum_i \theta_i u_i \ln u_i \geqq (\sum \theta_j u_j) \ln \sum \theta_i u_i - (\sum_j \theta_j u_j) \ln (\sum_k \theta_k) \geqq 0$ which follows using $\sum \theta_i u_i = 1$, $\sum_k \theta_k = 1 + \hat{\eta} \sum a_i p_i \leqq 1$ from $\mathbf{p} \in \Lambda$ and (3.2). Thus $\hat{p}_i = X_i/[n(1 + \hat{\eta} a_i)]$ minimizes and satisfies the constraints.

Finally for case (3.1 c) again assume $\mathbf{p} \in \Lambda$. Then

$$\hat{p}_1 = \sum_{i=2}^{K} X_i a_i/[n(a_i - a_1)] \,, \qquad \hat{p}_i = X_i a_1/[n(a_1 - a_i)] \qquad i = 2, \cdots, K$$

minimize since $a_*(X) \geqq 0$, $a_1 < 0$ gives $X_1 = 0$ and

$$\sum_{i=1}^{K} \frac{X_i}{n} \ln \frac{X_i}{np_i} = \sum_{i=2}^{K} \frac{X_i}{n} \ln \left(\frac{a_1 - a_i}{a_1}\right) + \sum_{i=2}^{K} \theta_i u_i \ln u_i \tag{3.5}$$

where $\theta_i = p_i(a_1 - a_i)/a_1$, $u_i = X_i a_1/[np_i(a_1 - a_i)]$. Using (3.4) again, $\sum_i \theta_i u_i \ln u_i \geqq 0$ by a similar argument since $\sum \theta_i u_i = 1$, $\sum_k \theta_k = \sum_i p_i(a_1 - a_i)/a_1 = 1 - \sum a_i p_i/a_1 \leqq 1$. Verifying $\sum_{i=1}^K a_i \hat{p}_i = 0, \hat{p}_i \geqq 0, \sum \hat{p}_i = 1$ completes the proof.

4. Large sample theory for contiguous alternatives. Let $\mathbf{p}^{(0)} \in \Omega_0$ satisfy $\sum_i a_i p_i^{(0)} = 0$ and consider a sequence of nearby or contiguous alternatives $\mathbf{p}^{(n)} \to \mathbf{p}^{(0)}$ where $\sum_i a_i p_i^{(n)} \geqq 0$ and $p_i^{(n)} = p_i^{(0)} + \xi_i n^{-\frac{1}{2}}$. Here $\sum_i \xi_i = 0$ and $\sum_i a_i \xi_i \geqq 0$. Next define

$$Y_i^{(n)} = (X_i - np_i^{(n)})/(np_i^{(n)})^{\frac{1}{2}} \tag{4.1}$$

so that the random vector $\mathbf{Y}^{(n)} = (Y_1^{(n)}, Y_2^{(n)}, \cdots, Y_K^{(n)}) \to_{\mathscr{L}} \mathscr{N}_K(0, \Sigma)$ as $n \to \infty$. The $K \times K$ covariance matrix of rank $K - 1$ is given by $\Sigma = (\delta_{ij} - (p_i^{(0)} p_j^{(0)})^{\frac{1}{2}})$ where the Kroneker $\delta_{ij} = 1$, if $i = j$, and 0 otherwise. The following theorem holds:

THEOREM 2. *The limiting distribution of the likelihood ratio statistic for the sequence of alternatives* $\mathbf{p}^{(n)}$ *is given by*

$$2nI(\mathbf{X}/n, \Lambda) \to_{\mathscr{L}} [\max(0, z + \delta)]^2 \tag{4.2}$$

where $z: N(0, 1)$ *and* $\delta = \sum_i a_i \xi_i^{(0)}/(\sum_j a_j^2 p_j^{(0)})^{\frac{1}{2}}$.

PROOF. Define

$$\begin{aligned} \tilde{\eta}_n(\mathbf{X}) &= \hat{\eta}_n(\mathbf{X}) \qquad \text{if } \textstyle\sum a_i X_i/n > 0\,, \quad a_*(\mathbf{X}) < 0 \\ &= 0 \qquad\qquad \text{otherwise.} \end{aligned}$$

Similarly define

$$\tilde{I}_n = \sum (X_i/n) \ln (1 + \tilde{\eta}_n a_i)\,.$$

We have for $\varepsilon > 0$

$$P[|2nI_n - 2n\tilde{I}_n| > \varepsilon] \leqq P[\textstyle\sum a_i X_i/n > 0, a_*(\mathbf{X}) \geqq 0] \to 0$$

as $n \to \infty$ using the consistency of $\mathbf{X}/n$ and $p_j^{(0)} > 0$, $a_1 < 0$. Thus it remains to show (4.2) for $\tilde{I}_n$. From (3.2) it follows that

$$\hat{\eta}_n = (\textstyle\sum_i a_i X_i/n)/\sum_j a_j^2 X_j/n(1 + \hat{\eta}_n a_j) \tag{4.3}$$

using $(1 + \hat{\eta}_n a_i)^{-1} = 1 - \hat{\eta}_n a_i + (\hat{\eta}_n a_i)^2 (1 + \hat{\eta}_n a_i)^{-1}$, $\hat{\eta}_n > 0$. Thus

$$\begin{aligned} n^{\frac{1}{2}}\tilde{\eta}_n &= \max[0, (n^{\frac{1}{2}} \textstyle\sum_i a_i X_i/n)/\sum_j a_j^2 X_j/n(1 + \tilde{\eta}_n a_j)] \\ &\to_{\mathscr{L}} \max[0, \textstyle\sum_i a_i (Y_i (p_i^{(0)})^{\frac{1}{2}} + \xi_i)/\sum_j a_j^2 p_j^{(0)}] \\ &= \max[0, (z + \delta)/(\textstyle\sum a_j^2 p_j^{(0)})^{\frac{1}{2}}] \end{aligned} \tag{4.4}$$

where $z = \sum a_i Y_i (p_i^{(0)})^{\frac{1}{2}}/(\sum a_j^2 p_j^{(0)})^{\frac{1}{2}} : N(0, 1)$, $Y_i^{(n)} \to_{\mathscr{L}} Y_i$, $i = 1, \cdots, K$. Using a limited expansion for the log,

$$\begin{aligned} 2n\tilde{I}_n &= 2n \textstyle\sum_i (X_i/n) \ln (1 + \tilde{\eta}_n a_i) \\ &= 2n \textstyle\sum_i (X_i/n) \left[\tilde{\eta}_n a_i - \left(\frac{\tilde{\eta}_n a_i}{2} \right)^2 (1 + o_p(1)) \right] \\ &\to_{\mathscr{L}} [\max(0, z + \delta)]^2\,. \end{aligned}$$

For the ad hoc statistic (2.2) the limiting distribution is that of $z + \delta$ so that for $\alpha \leqq \frac{1}{2}$ the tests are asymptotically equivalent under contiguous alternatives.

5. Large deviation efficiency comparisons. We consider a fixed alternative $\mathbf{p} \in \Omega_0 - \Lambda$ and contrast the likelihood ratio statistic and the ad hoc statistic by comparing exponential rates of type I error convergence to zero ($\lim_n n^{-1} \ln \alpha_n$) while equating type II errors $\beta_n \to \beta$, $0 < \beta < 1$. For the likelihood ratio test it can be shown that

$$\lim n^{-1} \ln \alpha_n = -\sum_i p_i \ln (1 + \eta a_i) = -d < 0 \tag{5.1}$$

where $0 < \eta < -1/a_1$ satisfies $\sum_i p_i/(1 + \eta a_i) = 1$. Similarly for the ad hoc statistic, we will show

$$\lim n^{-1} \ln \alpha_n = -I(A(c), \Lambda) \tag{5.2}$$

where

$$A(c) = \{\boldsymbol{\nu} \mid \sum_i a_i \nu_i \geqq c(\sum a_i^2 \nu_i)^{\frac{1}{2}},\ \sum \nu_i = 1,\ \nu_i \geqq 0\} . \tag{5.3}$$

The large deviation efficiency is then given by the ratio of (5.2) to (5.1) and is equivalent to the exact Bahadur efficiency.

To show (5.2) consider the critical region for the ad hoc statistic:

$$A^{(n)}(c_n) = \{\mathbf{x}/n \mid \sum a_i x_i/n \geqq c_n(\sum a_i^2 x_i/n)^{\frac{1}{2}},\ \sum x_i/n = 1,\ \text{integer } x_i \geqq 0\} .$$

For $\varepsilon > 0$ and n sufficiently large we have

$$A^{(n)}(c + \varepsilon) \subset A^{(n)}(c_n) \subset A^{(n)}(c - \varepsilon) \tag{5.4}$$

where the sequence of critical values c_n satisfies

$$c_n \to c = \sum_i a_i p_i/(\sum a_j^2 p_j)^{\frac{1}{2}} \tag{5.5}$$

using consistency, $\mathbf{X}/n \to_{\mathscr{L}} \mathbf{p}$, and $\beta_n \to \beta$, $0 < \beta < 1$.

We next show that $A^{(n)}(c)$ is regular relative to Λ:

$$\lim_n I(A^{(n)}(c), \Lambda) = I(A(c), \Lambda) . \tag{5.6}$$

Using Lemma 4.5 of Hoeffding (1965) there exists a point $\boldsymbol{\nu}^*$ on the boundary of the closure of $A(c)$ and a sequence $\boldsymbol{\nu}^{(n)} \in A^{(n)}(c)$ such that $\boldsymbol{\nu}^{(n)} \to \boldsymbol{\nu}^*$,

$$I(\boldsymbol{\nu}^*, \Lambda) = I(A(c), \Lambda) \leqq I(A^{(n)}(c), \Lambda) \leqq I(\boldsymbol{\nu}^{(n)}, \Lambda) . \tag{5.7}$$

By Wijsman's lemma (Hoeffding (1965), Lemma 4.4b), since Λ is the closure of a subset of Ω_0, it follows that $I(\cdot, \Lambda)$ is continuous. Thus

$$\lim_n I(\boldsymbol{\nu}^{(n)}, \Lambda) = I(\boldsymbol{\nu}^*, \Lambda) = I(A(c), \Lambda) . \tag{5.8}$$

Regularity holds since (5.6) follows from (5.7) and (5.8). By Theorem 6.1 of Hoeffding and (5.4) the expression

$$\lim_n n^{-1} \ln \alpha_n = \lim_n \sup \{n^{-1} \ln P(A^{(n)}(c_n) \mid \mathbf{p}) \mid p \in \Lambda\}$$

is bounded above and below by $-I(A^{(n)}(c - \varepsilon), \Lambda)$ and $-I(A^{(n)}(c + \varepsilon), \Lambda)$. Using regularity, these bounds approach $-I(A(c-\varepsilon)\ \Lambda)$ and $-I(A(c+\varepsilon), \Lambda)$. Letting $\varepsilon \downarrow 0$, (5.2) follows using continuity of $I(A(c), \Lambda)$ as a function of c.

The limit (5.1) can be similarly established with the constant d obtained by minimizing the corresponding information quantity. The large deviation efficiency (also Bahadur efficiency)

$$e(\mathbf{p}) = I(A(c), \Lambda)/d \tag{5.9}$$

where c is given by (5.5), $A(c)$ by (5.3), Λ by the hypothesis, and d by (5.1) satisfies $e(\mathbf{p}) \leqq 1$ since $\boldsymbol{\nu} = \mathbf{p} \in A(c)$. With some exceptions (e.g., $K = 2$ or cases with only two distinct values of a_i) we typically have $e(\mathbf{p}) < 1$ as is expected from the results of Brown (1971). For an illustration, consider the case of $K = 3$ and $\mathbf{a} = (-1, 0, 1)$. Then evaluating (5.1) we obtain $\eta = (p_3 - p_1)/(1 - p_2)$ $d = p_1 \ln(2p_1/(p_1 + p_3)) + p_3 \ln(2p_3/(p_1 + p_3))$. Evaluating (5.2) gives

$$I(A(c), \Lambda) = [(1-c)\ln(1-c) + (1+c)\ln(1+c)]/2$$

where $c = (p_3 - p_1)/(p_1 + p_3)^{\frac{1}{2}}$. Table 1 gives efficiencies for various $\mathbf{p} \in \Omega_0 - \Lambda$ along the line $p_3 = 1 - 2p_1$, $p_1 \leqq \frac{1}{3}$. It should be noted that for this example a UMP unbiased test exists which rejects for large values of X_3 given X_2, $X_1 + X_3$ and has a null binomial distribution.

TABLE 1
Efficiency values for selected $\mathbf{p}$, $\mathbf{a} = (-1, 0, 1)$, $p_3 = 1 - 2p_1$

p_1	0	.1	.2	.3	$\frac{1}{3}$
p_2	0	.1	.2	.3	$\frac{1}{3}$
p_3	1	.8	.6	.4	$\frac{1}{3}$
c	1	.7379	.4472	.1195	0
d	ln 2	.3099	.1046	.007167	0
$I(A(c), \Lambda)$	ln 2	.3047	.1036	.007160	0
$e(\mathbf{p})$	1	.9834	.9903	.9990	1

6. An application to disease screening. Consider the problem of comparing two different disease screening procedures such as mammography and a history questionnaire for breast cancer (Lundgren and Jakobsson (1976)). Table 2

TABLE 2
Mammography vs. *history for breast cancer screening*

	True neg. ($k = 0$) / True pos. ($k = 1$)	History Questionnaire negative ($j = 0$)	History Questionnaire positive ($j = 1$)	Total
Mammography	negative ($i = 0$)	5739	65 / 0	5804
	positive ($i = 1$)	168 / 29	8 / 5	210
	Total	5936	78	6014

summarizes results for 6014 women screened at Sandviken Sweden using both procedures.

Ignoring the earlier cell order, denote $\mathbf{X} = (5739, 65, 0, 168, 29, 8, 5)$ and the corresponding vector $\mathbf{p} = (p_{00\cdot}, p_{010}, p_{011}, p_{100}, p_{101}, p_{110}, p_{111})$ where X_{ijk} and p_{ijk} have subscripts $i, j, k = 0$ or 1 according as the mammography, history, and true (histologically confirmed) results are negative or positive. Here $(\cdot)$ indicates summation ($p_{00\cdot} = p_{000} + p_{001}$ because of lack of follow-up when both screenings are negative).

For purposes of illustration assign costs as follows:

\$2: cost of history administration
\$25: cost of mammography
\$200: cost of follow-up for a false positive (false alarm)
\$10,000: cost of a false negative (true cancer not detected).

Then the unit differences in expected costs can be expressed in the form (1.2) (history minus mammography)

$$(2 + 200p_{1\cdot 0} + 10000p_{0\cdot 1}) - (25 + 200p_{\cdot 10} + 10000p_{\cdot 01}) = \mathbf{ap}' \tag{6.1}$$

where $\mathbf{a} = (-23, -223, 9977, 177, -10023, -23, -23)$. The hypothesis of preference for mammography (H_0: $\mathbf{ap}' \leqq 0$) is rejected in favor of preference for the history (H_1: $\mathbf{ap}' > 0$) using both statistics. The ad hoc statistic gives a value of 2.434 with corresponding significance level $\hat{\alpha} = .0075$ using the normal approximation. The likelihood ratio statistic gives $2nI(\mathbf{X}/n, \Lambda) = 8.628$ with significance level $\hat{\alpha} = .0017$ using a normal approximation on $(2nI(\mathbf{X}/n, \Lambda))^{\frac{1}{2}}$.

Different cost assignments can change significance probabilities. For example, an additional cost for mammography due to an economic assessment of radiation exposure would result in an even more significant result.

Acknowledgments. Thanks go to Dr. Judith Ladinsky for bringing this problem to my attention. Thanks also go to Stephen Robinson for nonlinear programming assistance and to the referee for helpful suggestions and corrections.

REFERENCES

[1] Bahadur, R. R. (1971). *Some Limit Theorems in Statistics*. Regional Conference Series in Applied Mathematics **4**. SIAM, Philadelphia.

[2] Brown, L. D. (1971). Nonlocal optimality of appropriate likelihood ratio tests. *Ann. Math. Statist.* **42** 1206–1240.

[3] Hoeffding, W. (1965). Asymptotically optimal tests for multinomial distributions. *Ann. Math. Statist.* **36** 369–408.

[4] Lundgren, B. and Jakobsson, S. (1976). Single view mammography—a simple and efficient approach to breast cancer screening. *Cancer* **38** 1124–1129.

[5] Mangasarian, O. L. (1969). *Nonlinear Programming*. McGraw-Hill, New York.

Department of Statistics
University of Wisconsin
Madison, Wisconsin 53706