Ann. Inst. Statist. Math.
Vol. 42, No. 1, 133–147 (1990)

SMOOTHING OF LIKELIHOOD RATIO STATISTIC FOR EQUIPROBABLE MULTINOMIAL GOODNESS-OF-FIT

BALAKRISHNA S. HOSMANE*

Division of Statistics, Northern Illinois University, DeKalb, IL 60115, U.S.A.

(Received September 28, 1987; revised March 10, 1989)

Abstract. The likelihood ratio chi-square criterion for testing goodness-of-fit in k cell multinomials is known to overestimate significance for small and moderate sample sizes (see, e.g., Larntz (1978)). Therefore, the usual chi-square approximation to the upper tail of the likelihood ratio statistic G^2, is not satisfactory. Several authors have derived adjustments (e.g., Williams (1976), Smith *et al.* (1981), Hosmane (1987*b*)), so that the asymptotic mean of G^2 matches the mean of the asymptotic chi-square distribution in the hope that the distribution of G^2 would improve. In this paper, a new adjustment to G^2 is determined on the basis of the n^{-1}-order term (n being the total number) of the Edgeworth expansion of the distribution of smoothed G^2. Monte Carlo results indicate that the modified G^2 outperforms the unadjusted G^2.

Key words and phrases: Likelihood ratio statistic, Edgeworth expansion.

1. Introduction

In this paper, we consider smoothing of the likelihood ratio statistic G^2, for testing the equiprobability hypothesis in a k cell multinomial distribution. Since equiprobable class intervals produce the most sensitive tests (see, e.g., Cohen and Sackrowitz (1975), Spruill (1977) and Bednarski and Ledwina (1978)), several authors have considered small sample studies of the distribution of G^2 in this case. Chapman (1976), Larntz (1978), Koehler and Larntz (1980) and Lawal (1984) examined the error in approximating the distribution of G^2 with a chi-square distribution. Good *et al.* (1970) computed the least squares fit to $\log_{10} P^*(G^2 > a)$, where $P^*(G^2 > a)$ is a smoothed version of the exact tail probability of G^2.

It is known that the likelihood ratio chi-squared criterion overestimates significance, in the sense that the null hypothesis H_0 is rejected too often in relation to the nominal level of significance when H_0 is true for

*Present address: Abbott Laboratories, Department 436, Abbott Park, IL 60064, U.S.A.

moderate sample sizes (see, e.g., Larntz (1978), Smith *et al.* (1981), Lawal (1984) and Hosmane (1987*b*)). Williams (1976), Smith *et al.* (1981) and Hosmane (1987*b*) have suggested downward multiplicative correction to G^2 for testing goodness-of-fit of multinomial distributions. Larntz (1978) studied the behavior of G^2, concluding that the aberrant behavior of G^2 is due to very small observed counts in the table. The classical *ad hoc* procedure of adding 0.5 to all frequencies can be used, but it does not perform well. In this paper, we consider the following smoothing procedure:

C2: Add a positive constant to all the frequencies in the table.

This adjustment procedure has been considered by Hosmane (1986, 1987*a* and 1987*b*) to smooth statistics in the analysis of categorical data.

In this paper, we determine the suitable smoothing constants in the adjustment procedure C2 on the basis of the n^{-1}-order term of the Edgeworth expansion of the distribution of smoothed G^2. Here we assume the distribution of smoothed G^2 follows fairly closely to a smooth curve for small-to-moderate samples (see, e.g., Smith *et al.* (1981)). The analytical study involving asymptotic considerations was combined with Monte Carlo studies for finite samples. Some relevant findings of these studies are reported in Section 3.

2. Modified likelihood ratio statistic

Let $Y = (Y_1, \ldots, Y_k)'$ be a multinomial random vector with parameters (π, n) where $\pi = (\pi_1, \ldots, \pi_k)'$, $0 < \pi_j < 1$ for all j, $\sum_{j=1}^{k} \pi_j = 1$ and $n = \sum_{j=1}^{k} Y_j$. Consider a one-way classification, for a single factor B with k (fixed) levels, with observed frequencies Y_j, $j = 1, \ldots, k$. The null hypothesis is expressed as

$$H_0 \colon \pi_j = p_j, \quad j = 1, \ldots, k, \tag{2.1}$$

where p_j is the hypothesised probability of the j-th cell. In this paper, we consider the equiprobable hypothesis where $p_j = 1/k$ for all j. The likelihood ratio statistic G^2 to test H_0 is

$$G^2 = 2 \sum_{j=1}^{k} Y_j \ln (Y_j / np_j) . \tag{2.2}$$

It is known that G^2 has an asymptotic chi-square distribution with $(k - 1)$ degrees of freedom (df) when H_0 is true.

Now, using the C2 procedure we obtain the modified G^2, say $\tilde{G}^2(Y)$ given by

$$\tilde{G}^2(Y) = 2\left(1 + \frac{c}{n + a_+}\right)\sum_{j=1}^{k}(Y_j + a_j)\ln\{(Y_j + a_j)/(n + a_+)p_j\}, \tag{2.3}$$

where $a_+ = \sum_j a_j$, a_i is a non-negative constant such that $Y_j + a_j > 0$ for all j with probability one and $(1 + c/(n + a_+))$ is similar to the Bartlett adjustment factor (see Barndorff-Nielsen and Cox (1984)).

We now derive the Edgeworth expansion of the distribution of $\tilde{G}^2(Y)$ using the fact that the distribution of the likelihood ratio statistic in this situation follows fairly closely to a smooth curve for small-to-moderate samples (see, e.g., Smith *et al.* (1981)). In order to justify the Edgeworth expansion of the distribution of $\tilde{G}^2(Y)$ given in the next section, we have to exclude cases where the lattice character of the statistic is too pronounced (see, e.g., Knusel and Michalk (1987)).

THEOREM 2.1. *The Edgeworth expansion for the distribution of the statistic $\tilde{G}^2(Y)$ can be expressed as*

$$\begin{aligned} P(\tilde{G}^2(Y) < d) = P(\chi^2_{k-1} < d) \\ + \frac{1}{12n}\Bigg\{ P(\chi^2_{k-1} < d)\Bigg[(1 - S) + 6\Bigg(2\sum_{j=1}^{k}\frac{a_j}{p_j} \\ - \sum_{j=1}^{k}\frac{a_j^2}{p_j} + a_+^2 - (1 + k)a_+ - c(k - 1)\Bigg)\Bigg] \\ + P(\chi^2_{k+1} < d)\Bigg[(S - 1) + 6\Bigg(\sum_{j=1}^{k}\frac{a_j^2}{p_j} \\ - \sum_{j=1}^{k}\frac{a_j}{p_j} - a_+^2 + a_+ + c(k - 1)\Bigg)\Bigg] \\ - 6P(\chi^2_{k+3} < d)\Bigg[\sum_{j=1}^{k}\frac{a_j}{p_j} - ka_+\Bigg]\Bigg\} + O(n^{-3/2}) \end{aligned} \tag{2.4}$$

where χ^2_v represents a chi-square random variable with v degrees of freedom and $S = \sum_{j=1}^{k} p_j^{-1}$. The proof of the above theorem is given in the Appendix.

Here we note that for $a_j = 0$ for all j and $c = 0$ in (2.4), we obtain the result (4.8) obtained by Siotani and Fujikoshi (1984). In practice, we add the same constant, say a, to every cell i.e., $a_j = a$ for all j to smooth G^2. Then (2.4) reduces to

$$(2.5)\quad P(\tilde{G}^2(Y) < d) = P(\chi^2_{k-1} < d) + \frac{1}{12n}\{P(\chi^2_{k-1} < d)[(1-S) + 6(Sa(2-a) + ka(ka-k-1) - c(k-1))] + P(\chi^2_{k+1} < d)[S - 1 + 6(aS(a-1) - ka(ka-1) + c(k-1))] - 6P(\chi^2_{k+3} < d)a(S-k^2)\} + O(n^{-3/2}).$$

For $a = 0$, n^{-1}-order term in (2.5) is eliminated by selecting

$$(2.6)\qquad c = \frac{(1-S)}{6(k-1)}.$$

This c, corresponding to $a = 0$, is exactly the one obtained using moment correction to G^2 (see, e.g., Hosmane (1987*b*)). Also, Williams's (1976) correction q^{-1} in this situation is given by

$$q^{-1} = \left(1 + \frac{S-1}{6n(k-1)}\right)^{-1} = 1 + \frac{c}{n} + O(n^{-3/2}),$$

which agrees with our correction up to the n^{-1}-order term.

In the case of the equiprobability hypothesis, we have $S = k^2$ and hence (2.5) reduces to

$$(2.7)\quad P(\tilde{G}^2(Y) < d) = P(\chi^2_{k-1} < d) + \frac{1}{12n}[P(\chi^2_{k+1} < d) - P(\chi^2_{k-1} < d)] \cdot [(k^2-1) + 6(ka - k^2a + c(k-1))] + O(n^{-3/2}).$$

Then the choice of c for fixed a, in order to eliminate the n^{-1}-order term in (2.7) is

$$(2.8)\qquad c = \frac{1}{6}(6ak - k - 1),$$

and (2.7) reduces to

$$P(\tilde{G}^2(Y) < d) = P(\chi^2_{k-1} < d) + O(n^{-3/2}).$$

This choice of c corresponding to $a = 0$ for testing the equiprobability hypothesis is the same as the one obtained using moment correction to G^2

(see, e.g., Hosmane (1987*b*)). This indicates that improving the asymptotic mean of G^2 improves the approximate chi-square distribution as noted earlier by Lawley (1956), Williams (1976) and Barndorff-Nielsen and Cox (1984).

In the case of the asymmetric hypothesis (i.e., $p_j \neq 1/k$), the moment correction to G^2 for fixed a is given by

$$c = \frac{1}{k-1}\left\{ a(S-k) - a^2(S-k^2) + \frac{(1-S)}{6} \right\} \tag{2.9}$$

(see Hosmane (1987*b*)). For this choice of c, the equation (2.5) reduces to

$$P(\tilde{G}^2(Y) < d) = \frac{a}{2n}\{SP(\chi^2_{k-1} < d) + (k^2 - S)P(\chi^2_{k+3} < d)\} + O(n^{-3/2}) .$$

Therefore, for $a \neq 0$, improving the asymptotic mean of G^2 does not necessarily improve the asymptotic chi-square distribution for testing the asymmetric hypothesis.

3. Monte Carlo investigation

A Monte Carlo simulation study was carried out to assess the relative performance of smoothed G^2 statistic in finite samples. We considered 225 tables of sizes varying from 2 to 10 (i.e., $k = 2(1)10$) with $n/k = 1(1)25$ satisfying the equiprobability hypothesis H_0. For each null hypothesis, multinomial random variables Y_i with probability $1/k$ and sample size n are generated using an IMSL (1987) subroutine. Independent random samples were obtained 1,000 times for each distribution. For each simulation the adjusted G^2 were computed, using the C2 procedure, with different combinations of (a, c) values, for testing the equiprobability null hypothesis H_0. The empirical levels of significance, $\hat{\alpha}$, attained were computed as the proportion of times the value of G^2 exceeded the asymptotic critical value of χ^2_α for the nominal value $\alpha = 0.05$ and $\alpha = 0.01$ with $(k-1)$ degrees of freedom.

The C2 procedure was studied for the combination (a, c) with $a = 0.1$, 0.2, 0.25, 0.3, 0.4 and 0.5 where c is the optimal value as indicated in Section 2. For convenience in labelling the tables, the various Monte Carlo setups, each consisting of one value of a together with the value of c, given in the previous section, are labelled (1) through (6) as follows:

$$\tilde{G}^2\text{: (1) 0.1,}\quad \text{(2) 0.2,}\quad \text{(3) 0.25,}\quad \text{(4) 0.3,}\quad \text{(5) 0.4,}\quad \text{(6) 0.5} .$$

We also considered modified G^2 with $a = 0$ and $c = (1 - S)/6(k-1)$

denoted by G_0^2; $a = 0.5$ and $c = 0$ denoted by $G_{0.5}^2$. For the sake of comparison, the tables report values of $\hat{\alpha}$ for the unadjusted G^2 denoted by G_u^2. Also, the adjusted G^2 suggested by Williams (1976) and Smith *et al.* (1981) are included in our study; these are denoted by G_W^2 and G_{SRMS}^2, respectively.

All computer program were double precision and written in FORTRAN. The values of $\hat{\alpha}$ are given in Table 1 for four cell and eight cell multinomials (i.e., $k = 4$ and $k = 8$) satisfying the equiprobability hypothesis with $n/k = 2$, 5 *and* 10. They typify the behavior observed for the other tables considered in our study. It is seen that if many expected cell frequencies are small, the modified G^2 statistics with $a > 0$ tend to underestimate α, and the underestimation of α seems to increase as a increases. As expected, unadjusted G^2 overestimates α and $G_{0.5}^2$ seems to underestimate α in most cases. For moderate to large expected cell frequencies, G_W^2,

Table 1. Empirical levels (× 1000) for G^2 statistics at $\alpha = 0.05$ and $\alpha = 0.01$[†] in four cell and eight cell multinomials for testing the equiprobability hypothesis.

Statistics	$k = 4$			$k = 8$		
	$n/k =$			$n/k =$		
	2	5	10	2	5	10
G_u^2	32	58	68	102	86	52
	(15)	(24)	(14)	(21)	(24)	(12)
G_W^2	32	52	66	62	79	47
	(7)	(19)	(14)	(8)	(22)	(8)
G_{SRMS}^2	32	52	66	69	81	47
	(7)	(19)	(14)	(10)	(22)	(8)
G_0^2	32	51	66	62	79	47
	(7)	(19)	(4)	(8)	(22)	(8)
(1)	32	52	66	46	74	47
	(0)	(16)	(14)	(5)	(20)	(8)
(2)	32	52	66	40	72	46
	(0)	(16)	(14)	(4)	(20)	(8)
(3)	32	52	66	38	71	45
	(0)	(15)	(14)	(4)	(20)	(8)
(4)	32	52	66	37	71	45
	(0)	(15)	(14)	(4)	(19)	(8)
(5)	32	52	66	33	67	45
	(0)	(15)	(14)	(4)	(18)	(8)
(6)	32	52	66	28	65	45
	(0)	(13)	(14)	(3)	(16)	(8)
$G_{0.5}^2$	4	44	71	8	33	47
	(2)	(1)	(16)	(1)	(5)	(4)

[†]The numbers in parentheses correspond to $\alpha = 0.01$.

G^2_{SRMS}, G^2_0 and $G^2_{(1)}$ statistics attained levels that are quite close to the nominal values.

In order to check whether $\hat{\alpha}$ is reasonably close to the nominal value α, we adopted Cochran's (1952) suggestion that $\hat{\alpha}$ should be between 4% and 6% at the nominal 5% level, and between 0.7% and 1.5% at the nominal 1% level. Table 2 gives the percentage of tables among the 225 tables which had acceptable $\hat{\alpha}$ values. It seems that G^2_0 and $G^2_{(1)}$ approximate α better than any other G^2 statistics considered in our study.

Table 2. Acceptabilities for G^2 statistics at $\alpha = 0.05$ and $\alpha = 0.01$ for testing the equiprobability hypothesis.

Statistics	Percentage of tables having acceptable $\hat{\alpha}$	
	5%	1%
G^2_u	62.68	67.11
G^2_{W}	69.33	72.00
G^2_{SRMS}	67.56	72.00
G^2_0	72.89	72.56
(1)	73.33	72.15
(2)	72.67	69.78
(3)	72.33	68.44
(4)	72.33	67.56
(5)	71.44	66.67
(6)	71.11	66.47
$G^2_{0.5}$	57.4	62.5

We would like to point out that although Smith *et al.* (1981) conclude that Williams' method is never more accurate than their method, our results indicate that G^2_{W} approximates α better than G^2_{SRMS} in most cases.

A similar study was also carried out to compare the estimated powers of G^2 statistics for $k = 2(1)10$ with $n/k = 1(1)25$ to test the equiprobability hypothesis H_0 against the following alternative hypothesis

$$H_a\colon \pi_i = \begin{cases} \dfrac{1}{k}\left[1 - \dfrac{\delta}{(k-1)}\right], & i = 1,\dots,k-1\,, \\[2ex] \dfrac{(1+\delta)}{k}\,, & i = k\,, \end{cases} \tag{3.1}$$

(see, Cressie and Read (1984)), where $-1 \le \delta \le k-1$ is fixed. In Tables 3 and 4, we report the empirical powers of G^2 statistics to test H_0 against H_a with $\delta = 0.5$ (i.e., a "bump" alternative) and $\delta = -0.9$ (i.e., a "dip" alternative) for $k = 4$ and $k = 8$, respectively. The results indicate that G^2_0, $G^2_{(1)}$, G^2_{W} and G^2_{SRMS} have similar powers. Also, for modified G^2, the power

Table 3. Empirical powers (× 1000) for G^2 statistics at $\alpha = 0.05$ and $\alpha = 0.01^{\dagger}$ in four cell multinomial for testing the equiprobability hypothesis.

Statistics	$k=4,\ \delta=0.5$			$k=8,\ \delta=-0.9$		
	$n/k=$			$n/k=$		
	2	5	10	2	5	10
G_u^2	57	157	281	126	677	992
	(31)	(54)	(108)	(68)	(594)	(940)
G_W^2	57	152	265	126	676	991
	(18)	(47)	(108)	(30)	(519)	(931)
G_{SRMS}^2	67	152	265	126	676	991
	(18)	(47)	(108)	(30)	(519)	(940)
G_0^2	57	147	265	126	676	991
	(2)	(47)	(108)	(30)	(519)	(931)
(1)	57	152	265	126	676	991
	(2)	(43)	(108)	(6)	(351)	(931)
(2)	57	152	265	126	676	991
	(2)	(43)	(108)	(6)	(351)	(929)
(3)	57	152	265	126	676	991
	(2)	(40)	(108)	(6)	(286)	(929)
(4)	57	152	265	126	676	991
	(2)	(40)	(108)	(6)	(286)	(919)
(5)	57	152	265	126	676	991
	(2)	(40)	(108)	(6)	(286)	(919)
(6)	57	152	265	126	676	991
	(2)	(40)	(108)	(6)	(255)	(919)
$G_{0.5}^2$	20	125	249	22	664	985
	(7)	(31)	(115)	(5)	(141)	(900)

†The numbers in parentheses correspond to $\alpha = 0.01$.

decreases as a increases, whereas $G_{0.5}^2$ seems to have low power, as expected, since it underestimates α in the null case.

Summarizing the above results, the statistics that perform best are G_0^2 and $G_{(1)}^2$, as these approximate α correctly in most cases by correcting the known tendency of unadjusted G^2 to overestimate significance and have similar powers.

4. Conclusion

Results in the preceding section suggest that the "usual" adjustment procedure of adding 0.5 to every cell does not perform well at all. Therefore, we do not recommend the "usual" procedure in testing the equiprobability hypothesis using the likelihood ratio statistic. The modified likelihood ratio statistics G_0^2 and $G_{(1)}^2$ are seen to be satisfactory in attaining

Table 4. Empirical powers (× 1000) for G^2 statistics at $\alpha = 0.05$ and $\alpha = 0.01$[+] in eight cell multinomial for testing the equiprobability hypothesis.

Statistics	$k = 8,\ \delta = 0.5$			$k = 8,\ \delta = -0.9$		
	$n/k =$			$n/k =$		
	2	5	10	2	5	10
G_u^2	122	141	175	192	541	897
	(24)	(40)	(66)	(36)	(173)	(639)
G_W^2	66	125	164	120	497	883
	(11)	(34)	(62)	(16)	(146)	(617)
G_{SRMS}^2	81	128	166	138	509	885
	(14)	(35)	(62)	(19)	(151)	(620)
G_0^2	66	125	164	120	497	882
	(11)	(34)	(62)	(16)	(146)	(617)
(1)	42	121	164	75	450	881
	(5)	(31)	(61)	(12)	(124)	(615)
(2)	38	116	164	63	421	881
	(5)	(30)	(61)	(9)	(111)	(609)
(3)	38	114	163	62	411	881
	(5)	(29)	(60)	(9)	(106)	(605)
(4)	37	113	163	57	405	880
	(5)	(28)	(60)	(9)	(100)	(604)
(5)	35	109	161	52	376	877
	(5)	(28)	(60)	(9)	(96)	(601)
(6)	33	105	161	44	355	874
	(4)	(28)	(60)	(7)	(90)	(596)
$G_{0.5}^2$	11	87	159	24	251	890
	(1)	(30)	(54)	(3)	(83)	(550)

[+]The numbers in parentheses correspond to $\alpha = 0.01$.

Type I error levels and to perform slightly better than the modified G^2 statistics suggested by Williams (1976) and Smith *et al.* (1981). However, the $G_{(1)}^2$ statistic has a slight edge over G_0^2 for moderate to large samples.

Acknowledgement

The author is grateful to the referee for pointing out a computational error in Section 3 and for helpful suggestions which greatly improved the paper.

Appendix

Define $W_j = n^{-1/2}(Y_j - np_j)$, $j = 1,\ldots,k$, and let $\boldsymbol{W} = (W_1,\ldots, W_r)'$ where $r = k - 1$. Then the random vector $\boldsymbol{W}$ is a lattice random vector which takes values in the lattice

$$L = \{\boldsymbol{w} = (w_1,\ldots, w_r)' : \boldsymbol{w} = n^{-1/2}(\boldsymbol{m} - n\boldsymbol{q}) \text{ and } \boldsymbol{m} \in M\}$$

where $\boldsymbol{q} = (p_1,\ldots,p_r)'$ and M is a set of integer vectors $\boldsymbol{m} = (m_1,\ldots, m_r)'$ such that $m_j \geq 0$ and $\sum_{j=1}^{r} m_j \leq n$. The asymptotic expansion of the probability mass function $P(\boldsymbol{W} = \boldsymbol{w})$ is given by the following lemma.

LEMMA A.1. (Siotani and Fujikoshi (1984)) *Let* $\boldsymbol{w} = n^{-1/2}(\boldsymbol{m} - n\boldsymbol{q})$. *If* $\boldsymbol{m} \in M$, *then*

$$\text{(A.1)} \qquad P(\boldsymbol{W} = \boldsymbol{w}) = n^{-r/2}\phi(\boldsymbol{w})\{1 + n^{-1/2}h_1(\boldsymbol{w}) + n^{-1}h_2(\boldsymbol{w}) + O(n^{-3/2})\}$$

where

$$\phi(\boldsymbol{w}) = (2\pi)^{-r/2}|\Omega|^{-1/2} \exp\left(-\frac{1}{2}\boldsymbol{w}'\Omega^{-1}\boldsymbol{w}\right)$$

is the multivariate Normal density function, and

$$h_1(\boldsymbol{w}) = -\frac{1}{2}\sum_{j=1}^{k}\frac{w_j}{p_j} + \frac{1}{6}\sum_{j=1}^{k}\frac{w_j^3}{p_j^2},$$

$$h_2(\boldsymbol{w}) = \frac{1}{2}[h_1(\boldsymbol{w})]^2 + \frac{1}{12}\left(1 - \sum_{j=1}^{k}\frac{1}{p_j}\right) + \frac{1}{4}\sum_{j=1}^{k}\frac{w_j^2}{p_j^2} - \frac{1}{12}\sum_{j=1}^{k}\frac{w_j^4}{p_j^3},$$

with $w_k = -\sum_{j=1}^{r} w_j$, $\Omega = \text{diagonal}\,(p_1,\ldots,p_r) - \boldsymbol{q}\boldsymbol{q}'$.

The above lemma gives a local Edgeworth approximation for the probability of $\boldsymbol{W}$ at each $\boldsymbol{w}$ in the lattice L. Hence the continuous multivariate Edgeworth approximation for the probability of any set D from (A.1) is

$$P(\boldsymbol{W} \in D) = \int\cdots\int_D \phi(\boldsymbol{w})\{1 + n^{-1/2}h_1(\boldsymbol{w}) + n^{-1}h_2(\boldsymbol{w})\}d\boldsymbol{w} + O(n^{-3/2})\,.$$

We now outline the proof of the theorem.

PROOF OF THE THEOREM 2.1. Let

$$D(d) = \{\boldsymbol{w} = (w_1,\ldots, w_r)' : \tilde{G}^2(\boldsymbol{Y}) < d\}\,,$$

where $w_k = -\sum_{j=1}^{r} w_j$. The distribution of $\tilde{G}^2(\boldsymbol{Y})$ can now be expressed as

$$P(\tilde{G}^2(\boldsymbol{Y}) < d) = P(\boldsymbol{W} \in D(d)) .$$

Consider the transformation

$$\text{(A.2)} \qquad \boldsymbol{z} = (z_1,\dots,z_r)' = H\boldsymbol{w} = B'\Delta^{-1/2}\begin{bmatrix} I_r \\ -1,\dots,-1 \end{bmatrix}\boldsymbol{w} ,$$

where I_r is the identity matrix of order r,

$$\Delta = \text{diagonal}\,(p_1,\dots,p_k) ,$$
$$\sqrt{\boldsymbol{p}} = (\sqrt{p_1},\dots,\sqrt{p_k}) ,$$

$B = (\boldsymbol{b}_1,\dots,\boldsymbol{b}_k)'$ is a $k \times r$ matrix such that $(B, \sqrt{\boldsymbol{p}})$ is an orthogonal matrix. Then noting that $H\Omega H' = I_r$ and $\sqrt{p_j}(\boldsymbol{b}_j'\boldsymbol{z}) = w_j$, we can express (A.1) as

$$\text{(A.3)} \qquad P(\boldsymbol{W} = \boldsymbol{w}) = n^{-r/2}|\Omega|^{-1/2}\{f(\boldsymbol{z}) + O(n^{-3/2})\} ,$$

where

$$\text{(A.4)} \qquad f(\boldsymbol{z}) = (2\pi)^{-r/2}\exp\left(-\frac{1}{2}\boldsymbol{z}'\boldsymbol{z}\right)\{1 + n^{-1/2}g_1(\boldsymbol{z}) + n^{-1}g_2(\boldsymbol{z})\}$$

with

$$g_1(\boldsymbol{z}) = -\frac{T_1}{2} + \frac{T_3}{6} ,$$
$$g_2(\boldsymbol{z}) = \frac{g_1^2(\boldsymbol{z})}{2} + \frac{(1-S)}{12} + \frac{T_2}{4} - \frac{T_4}{12}$$

and

$$T_1 = \sum_{j=1}^{k}\frac{\boldsymbol{b}_j'\boldsymbol{z}}{\sqrt{p_j}} , \qquad T_2 = \sum_{j=1}^{k}\frac{(\boldsymbol{b}_j'\boldsymbol{z})^2}{p_j} ,$$
$$T_3 = \sum_{j=1}^{k}\frac{(\boldsymbol{b}_j'\boldsymbol{z})^3}{\sqrt{p_j}} , \qquad T_4 = \sum_{j=1}^{k}\frac{(\boldsymbol{b}_j'\boldsymbol{z})^4}{p_j}$$

and

$$S = \sum_{j=1}^{k} p_j^{-1} .$$

By interpreting $f(z)$ as the continuous density function of the random vector Z, it is possible to approximate the distribution of the statistic $\tilde{G}^2(H^{-1}Z)$, assuming that the lattice character of $\tilde{G}^2(H^{-1}Z)$ is not too pronounced, as

$$P(\tilde{G}^2(Y) < d) = \int \cdots \int_{\tilde{D}(d)} f(z)\, dz \tag{A.5}$$

where

$$\tilde{D}(d) = \{z\colon z = Hw \text{ and } w \in D(d)\} . \tag{A.6}$$

Then the characteristic function of $\tilde{G}^2(H^{-1}Z)$ is defined by

$$\psi(t) = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \exp\{it\tilde{G}^2(H^{-1}z)\} f(z)\, dz . \tag{A.7}$$

We can express $\tilde{G}^2(H^{-1}z)$ using a Taylor series as

$$\begin{aligned} \tilde{G}^2(H^{-1}z) = z'z &+ n^{-1/2}\left\{ 2 \sum_{j=1}^{k} \frac{a_j(b_j'z)}{\sqrt{p_j}} - \frac{T_3}{3} \right\} \\ &+ n^{-1}\left\{ \sum_{j=1}^{k} \frac{a_j^2}{p_j} - \sum_{j=1}^{k} \frac{a_j(b_j'z)^2}{p_j} + \frac{T_4}{6} + cz'z - a_+^2 \right\} \\ &+ O(n^{-3/2}) . \end{aligned} \tag{A.8}$$

Since the asymptotic distribution of W, and hence Z, is multivariate normal, then

$$\frac{b_j'Z}{p_j\sqrt{n}} = O_p(n^{-1/2}) .$$

Also, since

$$\exp\{\alpha + n^{-1/2}\beta + n^{-1}v\} = e^{\alpha}\left\{ 1 + n^{-1/2}\beta + n^{-1}\left(v + \frac{\beta^2}{2} \right)\right\} + O(n^{-3/2}) ,$$

then using g_1 and g_2 from (A.4), $\psi(t)$ in (A.7) can be expressed as

(A.9) $$\psi(t) = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} (2\pi)^{-r/2} \exp\left\{ \frac{(2it-1)}{2} \mathbf{z}'\mathbf{z} \right\} \{1 + n^{-1/2}h_1(\mathbf{z}) + n^{-1}h_2(\mathbf{z})\} \cdot \{1 + n^{-1/2}g_1(\mathbf{z}) + n^{-1}g_2(\mathbf{z})\} d\mathbf{z} + O(n^{-3/2})$$

where

$$h_1(\mathbf{z}) = it\left[2Q_1 - \frac{T_3}{3} \right],$$

$$h_2(\mathbf{z}) = it\left[Q_0 - Q_2 + \frac{T_4}{6} + c\mathbf{z}'\mathbf{z} - a_+^2 \right] + 2[itQ_1]^2 + \frac{(itT_3)^2}{18} - \frac{2}{3}[itQ_1T_3]$$

with

$$Q_0 = \sum_{j=1}^{k} \frac{a_j^2}{p_j}, \quad Q_1 = \sum_{j=1}^{k} \frac{a_j(\mathbf{b}_j'\mathbf{z})}{\sqrt{p_j}} \quad \text{and} \quad Q_2 = \sum_{j=1}^{k} \frac{a_j(\mathbf{b}_j'\mathbf{z})^2}{p_j}.$$

Now substituting $\sigma^2 = (1 - 2it)^{-1}$, (A.9) reduces to

(A.10) $$\psi(t) = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} (2\pi)^{-r/2} \exp\left\{ -\frac{1}{2\sigma^2} \mathbf{z}'\mathbf{z} \right\} u(\mathbf{z}) d\mathbf{z} + O(n^{-3/2})$$

where

(A.11) $$\begin{aligned} u(\mathbf{z}) &= [1 + n^{-1/2}h_1(\mathbf{z}) + n^{-1}h_2(\mathbf{z})][1 + n^{-1/2}g_1(\mathbf{z}) + g_1(\mathbf{z}) + n^{-1}g_2(\mathbf{z})] \\ &= 1 + n^{-1/2}\left\{ 2itQ_1 - \frac{T_1}{2} + \frac{T_3}{6} - \frac{itT_3}{3} \right\} \\ &\quad + n^{-1}\left\{ \frac{1}{2}\left(\frac{T_3}{6} - \frac{T_1}{2} \right)^2 + \frac{(1-S)}{12} + \frac{T_2}{4} - \frac{T_4}{12} \right. \\ &\quad + itQ_0 - itQ_2 + \frac{itT_4}{6} + itc\mathbf{z}'\mathbf{z} \\ &\quad - ita_+^2 + 2(itQ_1)^2 + \frac{(itT_3)^2}{18} - \frac{itQ_1T_3}{3} \\ &\quad \left. - itT_1Q_1 + it\frac{T_1T_3}{6} - it\frac{T_3^2}{18} \right\}. \end{aligned}$$

From (A.10) it follows that

$$\text{(A.12)} \qquad \psi(t) = \sigma^r E[u(\boldsymbol{Z})] + O(n^{-3/2})$$

where $\boldsymbol{Z}$ is formally a normal random vector with mean $\mathbf{0}$ and covariance matrix $\sigma^2 I_r$. Now, using the moment formulas for a multivariate normal random vector $\boldsymbol{Z}$, and replacing $\boldsymbol{z}$ by $\boldsymbol{Z}$ in T_1, T_2, T_3, T_4, Q_1 and Q_2 of (A.11), we obtain

$$\begin{aligned}
&E(T_1) = 0, \quad E(T_3) = 0, \quad E(T_1^2) = \sigma^2(S - k^2), \\
&E(T_2) = \sigma^2(S - k), \quad E(T_1 T_3) = 3\sigma^4(S - k^2), \\
&E(T_4) = 3\sigma^4(S - 2k + 1), \quad E(T_3^2) = 3\sigma^6(5S - 3k^2 - 6k + 4), \\
\text{(A.13)} \quad &E(Q_1^2) = \sigma^2\left\{\sum_{j=1}^{k} \frac{a_j^2}{p_j} - a_+^2\right\}, \quad E(Q_2^2) = \sigma^2\left\{\sum_{j=1}^{k} \frac{a_j}{p_j} - a_+\right\}, \\
&E(Q_1 T_1) = \sigma^2\left\{\sum_{j=1}^{k} \frac{a_j}{p_j} - k a_+\right\} \quad \text{and} \\
&E(Q_1 T_3) = 3\sigma^4\left\{\sum_{j=1}^{k} \frac{a_j}{p_j} - k a_+\right\}.
\end{aligned}$$

Now from (A.12) and (A.13), we obtain

$$\begin{aligned}
\text{(A.14)} \quad \psi(t) = \sigma^r + \frac{\sigma^r}{12n}\Bigg\{&(1 - S) + 6\left[2\sum_{j=1}^{k} \frac{a_j}{p_j} - (1 + k)a_+ - c(k - 1)\right. \\
&\left. + a_+^2 - \sum_{j=1}^{k} \frac{a_j^2}{p_j}\right] + \sigma^2\left[(S - 1) + 6\left(\sum_{j=1}^{k} \frac{a_j^2}{p_j}\right.\right. \\
&\left.\left. - a_+^2 + c(k - 1) - \sum_{j=1}^{k} \frac{a_j}{p_j} + a_+\right)\right]\Bigg\} + O(n^{-3/2}).
\end{aligned}$$

Now, recall that $\sigma^r = (1 - 2it)^{-r/2}$, which is the characteristic function of a χ^2 random variable with $r = k - 1$ degrees of freedom. Inverting (A.14) and recalling that $\psi(t)$ is the characteristic function of the $\tilde{G}^2(\boldsymbol{Y})$, we obtain the result of the theorem.

References

Barndorff-Nielsen, O. E. and Cox, D. R. (1984). Bartlett adjustments to the likelihood ratio statistic and the distribution of the maximum likelihood estimator, *J. Roy. Statist. Soc. Ser. B*, **46**, 483–495.

Bednarski, T. and Ledwina, T. (1978). A note on biasedness of tests of fit, *Mathematische Operationsforschung und Statistik, Series Statistics*, **9**, 191–193.

Chapman, J. W. (1976). A comparison of the X^2, $-2 \log R$ and multinomial probability criteria for significance tests when expected frequencies are small, *J. Amer. Statist.*

Assoc., **71**, 854–863.

Cochran, W. G. (1952). The χ^2 test of goodness of fit, *Ann. Math. Statist.*, **23**, 315–345.

Cohen, A. and Sackrowitz, H. B. (1975). Unbiasedness of the chi-square, likelihood ratio and other goodness-of-fit tests for the equal cell case, *Ann. Statist.*, **3**, 959–964.

Cressie, N. A. and Read, T. R. C. (1984). Multinomial goodness-of-fit tests, *J. Roy. Statist. Soc. Ser. B*, **46**, 440–464.

Good, I. J., Gover, T. N. and Mitchell, G. J. (1970). Exact distributions for χ^2 and for the likelihood-ratio statistic for the equiprobable multinomial distribution, *J. Amer. Statist. Assoc.*, **65**, 267–283.

Hosmane, B. S. (1986). Improved likelihood ratio tests and Pearson chi-square tests for independence in two-dimensional contingency tables, *Comm. Statist. A—Theory Methods*, **15**, 1875–1888.

Hosmane, B. S. (1987*a*). On the adjusted Wald statistic for testing log-linear hypothesis in categorical data, *Sankhyā Ser. B*, **49**, 199–217.

Hosmane, B. S. (1987*b*). Improved likelihood ratio test for multinomial goodness of fit, *Comm. Statist. A—Theory Methods*, **16**, 3185–3198.

Knusel, L. and Michalk, J. (1987). Asymptotic expansion of the power function of the two-sample binomial test with and without randomization, *Metrika*, **34**, 31–44.

Koehler, K. J. and Larntz, K. (1980). An empirical investigation of goodness-of-fit statistics for sparse multinomials, *J. Amer. Statist. Assoc.*, **75**, 336–342.

Larntz, K. (1978). Small-sample comparisons of exact levels for chi-squared goodness-of-fit statistics, *J. Amer. Statist. Assoc.*, **73**, 253–263.

Lawal, H. B. (1984). Comparison of the X^2, Y^2, Freeman-Tukey and Williams' improved G^2 test statistics in small-samples of one-way multinomials, *Biometrika*, **71**, 415–418.

Lawley, D. N. (1956). A general method for approximating to the distribution of likelihood ratio criteria, *Biometrika*, **43**, 295–303.

Siotani, M. and Fujikoshi, Y. (1984). Asymptotic approximations for the distributions of multinomial goodness-of-fit statistics, *Hiroshima Math. J.*, **14**, 115–124.

Smith, P. J., Rae, D. S., Manderschield, R. W. and Silbergeld, S. (1981). Approximating the moments and distribution of the likelihood ratio statistic for multinomial goodness of fit, *J. Amer. Statist. Assoc.*, **76**, 737–740.

Spruill, M. C. (1977). Equally likely intervals in the chi-square test, *Sankhyā Ser. A*, **39**, 299–302.

Williams, D. A. (1976). Improved likelihood ratio tests for complete contingency tables, *Biometrika*, **63**, 33–37.