A POWER APPROXIMATION FOR THE MULTINOMIAL GOODNESS-OF-FIT TEST BASED ON A NORMALIZING TRANSFORMATION

Yuri Sekiya*, Nobuhiro Taneichi** and Hideyuki Imai***

Cressie and Read (1984) introduced the class of multinomial goodness-of-fit statistics, $\mathbf{R}^a$, based on measures of the divergence between discrete distributions. All $\mathbf{R}^a$ have the same chi-squared limiting null distribution. The power of $\mathbf{R}^a$ is usually approximated by a noncentral chi-squared distribution that is also the same for all a .

In this paper, we propose a new approximation of the power of $\mathbf{R}^a$. The new power approximation, NT, is a normal approximation based on a normalizing transformation. The NT approximation is numerically compared with the other approximations. As a result of the comparison, we find that the NT approximation is superior to the other approximations when $a=0$ (the loglikelihood ratio statistic) and is effective when $a<0$.

Key Words and Phrase: goodness of fit, multinomial distribution, normalizing transformation.

1. Introduction

Let $\mathbf{X}=(X_1, \dots, X_k)'$ be a random vector according to a multinomial distribution with parameters $n, \boldsymbol{\pi}=(\pi_1, \dots, \pi_k)'$ where $\sum_{j=1}^k X_j=n$, and π_j is the (true unknown) probability of an observation being classified into cell $E_j, j=1, \dots, k$. In order to test the null hypothesis $H_0: \boldsymbol{\pi}=\mathbf{p}$, for some completely specified probability vector $\mathbf{p}=(p_1, \dots, p_k)'$, Cressie and Read (1984) introduced the power divergence family of statistics

$$(1.1) \quad \mathbf{R}^a = 2nI^a\left(\frac{\mathbf{X}}{n}, \mathbf{p}\right),$$

where power divergence I^a is defined by

$$I^a(\mathbf{q}, \mathbf{p}) = \begin{cases} \frac{1}{a(a+1)} \sum_{i=1}^k q_i \left\{ \left(\frac{q_i}{p_i} \right)^a - 1 \right\} & (a \neq 0, -1) \\ \sum_{i=1}^k q_i \log \frac{q_i}{p_i} & (a=0) \\ \sum_{i=1}^k p_i \log \frac{p_i}{q_i} & (a=-1), \end{cases}$$

for the probability vectors $\mathbf{p}$ and $\mathbf{q}$. This class includes Pearson's X^2 statistic

Received December, 1997. Revised July, 1998. Accepted September, 1998.

* Kushiro Campus, Hokkaido University of Education, Kushiro, 085-8580 Japan.

** Obihiro University of Agriculture and Veterinary Medicine, Obihiro, 080-8555 Japan.

*** Division of Systems and Information Engineering, Hokkaido University, Sapporo, 060-8628 Japan.

$(X^2)(a=1)$, the loglikelihood ratio statistic $(LR)(a=0)$, the Freeman-Tukey statistic $(F^2)(a=-1/2)$, the modified loglikelihood ratio statistic $(GM^2)(a=-1)$, and the Neyman-modified X^2 statistic $(NM^2)(a=-2)$. Furthermore, $\mathbf{R}^{2/3}$ is the statistic recommended by Cressie and Read (1984) and Read and Cressie (1988). Under the null hypothesis H_0 , all $\mathbf{R}^a$ have the limiting distribution χ^2_{k-1} as $n \rightarrow \infty$ with the number of cells k being fixed (see Cressie and Read (1984)). Ordinary power approximation of $\mathbf{R}^a$ is based on the following large-sample results: for fixed k and local alternatives $H_{1,n}: \boldsymbol{\pi} = \mathbf{p} + \mathbf{c}/\sqrt{n}$, where $\mathbf{c} = (c_1, \dots, c_k)'$, and $\sum_{i=1}^k c_i = 0$, the limiting distribution of any $\mathbf{R}^a$ under $H_{1,n}$ is noncentral chi-square, $\chi^2_{k-1}(\delta)$, with the noncentrality parameter $\delta = \sum_{i=1}^k c_i^2/p_i$. Therefore, the power of $\mathbf{R}^a$ is evaluated to be the same for all a .

On the other hand, power approximations of $\mathbf{R}^a$ that vary with the statistic chosen and are effective for small samples have been studied. Broffitt and Randles (1977) showed that for fixed k and a nonlocal (fixed) alternative, a suitably normalized X^2 statistic has a limiting standard normal distribution. They proposed an approximation to the power of X^2 statistic against a nonlocal (fixed) alternative using the limiting distribution. Read and Cressie (1988) extended the normal approximation to other $\mathbf{R}^a (a > -1)$. Furthermore, Sekiya and Taneichi (1994) improved the normal approximation by correcting the first and second moments of $\mathbf{R}^a$ to higher order terms.

Drost et al. (1989) proposed two power approximations A^a and B^a . The A^a approximate distributions of $\mathbf{R}^a$ under alternatives as a linear combination of mutually independent noncentral chi-squared random variables. The B^a approximate the distributions as a linear expression of a single noncentral chi-squared random variable which has the same moment as A^a . From their small-sample calculations of many examples, they concluded that the A^a approximation is superior to the ordinary noncentral chi-squared approximation, the normal approximation, and the B^a approximation and is quite accurate for a wide range of a values. Though the A^a approximation is more accurate than the other approximations, it is difficult to compute. Moreover, it is supposed from Table 1 of Drost et al. (1989) that the A^a and B^a approximations become less accurate as the value of a decreases from 1.

In this paper, we consider a normal power approximation that is easier to compute. The approximation is obtained using a normalizing transformation of the statistics. In Section 2, we outline normal power approximations. In Section 3, a normal power approximation using a normalizing transformation is proposed. In Section 4, the approximation proposed in Section 3 is numerically compared with the other approximations.

2. Outline of normal power approximations

We consider a nonlocal (fixed) alternative

$$H_1: \boldsymbol{\pi} = \mathbf{q}, (\mathbf{q} \neq \mathbf{p}),$$

where $\mathbf{q} = (q_1, \dots, q_k)'$ is a completely specified probability vector which does not depend on n . If we put

$$(2.1) \quad T^a = I^a\left(\frac{\mathbf{X}}{n}, \mathbf{p}\right),$$

and

$$Y_i = \frac{X_i - nq_i}{\sqrt{n}}, \quad (i=1, \dots, k),$$

T^a is expanded under H_1 as follows:

$$(2.2) \quad T^a = G^a(\mathbf{Y}) + o_p\left(\frac{1}{n}\right),$$

where

$$G^a(\mathbf{Y}) = I^a(\mathbf{q}, \mathbf{p}) + \frac{1}{\sqrt{n}} \sum_{i=1}^k A_i^{(a)} Y_i + \frac{1}{2n} \sum_{i=1}^k B_i^{(a)} Y_i^2,$$

$$A_i^{(a)} = \begin{cases} \frac{1}{a} \left(\frac{q_i}{p_i} \right)^a & (a \neq 0) \\ \log \frac{q_i}{p_i} & (a = 0), \end{cases}$$

and

$$B_i^{(a)} = \frac{1}{q_i} \left(\frac{q_i}{p_i} \right)^a.$$

The mean and variance of T^a for $a > -1$ under H_1 are evaluated as

$$(2.3) \quad E(T^a) = \mu + \frac{1}{n} \mu_1 + o\left(\frac{1}{n\sqrt{n}}\right),$$

$$(2.4) \quad V(T^a) = \frac{1}{n} \mu_2 + o\left(\frac{1}{n\sqrt{n}}\right),$$

where

$$\mu = I^a(\mathbf{q}, \mathbf{p}),$$

$$\mu_1 = \frac{1}{2} \sum_{i=1}^k B_i^{(a)} M_{ii}^{(2)},$$

$$\mu_2 = \sum_{i=1}^k \sum_{j=1}^k A_i^{(a)} A_j^{(a)} M_{ij}^{(2)},$$

$$M_{ij}^{(2)} = \delta_{ij} q_i - q_i q_j,$$

and δ_{ij} denotes Kronecker's delta symbol. Let $\xrightarrow{L}$ denote convergence in distribution. Applying the multivariate central limit theorem to the distribution of $\mathbf{Y} = (Y_1, \dots, Y_k)'$, we obtain from (2.2), (2.3), and (2.4) that

$$(2.5) \quad \sqrt{\frac{n}{\mu_2}} (T^a - \mu) \xrightarrow{L} N(0, 1), \text{ as } n \rightarrow \infty,$$

under H_1 . Using (2.5), Read and Cressie (1988) derived a normal power approximation of $\mathbf{R}^a$ under H_1 . A normal power approximation for Pearson's

$X^2(a=0)$ proposed by Broffitt and Randles (1977) is nothing but an improvement of the normal power approximation by correcting the mean and variance of the statistic to higher order terms. Furthermore, Sekiya and Taneichi (1994) extended the improvement to other $R^a(a > -1)$ by correcting the mean and variance up to $O(1/n)$. We call this approximation the MCN approximation.

3. An improved normal power approximation using a normalizing transformation

In order to improve the normal power approximation given by (2.5), we consider a normalizing transformation of T^a . Konishi (1981) discussed a general method to find the normalizing transformation when we have already obtained the asymptotic expansion for the distribution of $f(T)$, where f is a monotone function of a statistic T . Similarly, we can find the normalizing transformation using the asymptotic expansion of the characteristic function or the cumulant generating function of $f(T)$. Goodness-of-fit statistics for the multinomial distribution are not distributed as a lattice distribution but as a complex discrete distribution. It is very difficult to obtain the asymptotic expansion for the distribution of $f(T^a)$, because we can not use the inversion formula. Therefore, we look for a normalizing transformation of T^a based on the cumulant generating function.

Let $f(T^a)$ be a strictly monotone function for $T^a \geq 0$, and twice continuously differentiable at $T^a = \mu$. If we define

$$(3.1) \quad g(T^a) = \sqrt{\frac{n}{\mu_2}} \left\{ \frac{1}{f'(\mu)} (f(T^a) - f(\mu)) - \frac{1}{n} \left(\mu_1 + \frac{\mu_2 f''(\mu)}{2f'(\mu)} \right) \right\},$$

$g(T^a)$ is also a strictly monotone function for $T^a \geq 0$, and twice continuously differentiable at $T^a = \mu$. The third moment about the mean of T^a for $a > -1$ under H_1 is evaluated as

$$(3.2) \quad E[\{T^a - E(T^a)\}^3] = \frac{1}{n^2} \mu_3 + o\left(\frac{1}{n^2 \sqrt{n}}\right),$$

where

$$\mu_3 = \sum_{i=1}^k \sum_{j=1}^k \sum_{l=1}^k (A_i^{(a)} A_j^{(a)} A_l^{(a)} M_{ijl}^{(3)} + 3A_i^{(a)} A_j^{(a)} B_l^{(a)} M_{il}^{(2)} M_{jl}^{(2)}),$$

and

$$M_{ijl}^{(3)} = \delta_{ij} \delta_{il} q_i - (\delta_{ij} q_i q_l + \delta_{il} q_i q_j + \delta_{jl} q_i q_j) + 2q_i q_j q_l.$$

Thus, from (2.3), (2.4), (3.1), and (3.2), the mean and first three moments about the mean of $g(T^a)$ for $a > -1$ under H_1 are evaluated as

$$\begin{aligned} E[g(T^a)] &= o\left(\frac{1}{n}\right), \\ V[g(T^a)] &= 1 + o\left(\frac{1}{\sqrt{n}}\right), \end{aligned}$$

and

$$E[\{g(T^a) - E(g(T^a))\}^3] = 3\sqrt{\frac{\mu_2}{n}} \left\{ \frac{f''(\mu)}{f'(\mu)} + \frac{\mu_3}{3\mu_2^2} \right\} + o\left(\frac{1}{n}\right),$$

respectively. Furthermore, the j -th ($j \geq 4$) cumulant of $g(T^a)$ under H_1 is $o(1/\sqrt{n})$. Therefore, if f satisfies the condition

$$(3.3) \quad \frac{f''(\mu)}{f'(\mu)} = -\frac{\mu_3}{3\mu_2^2},$$

$g(T^a)$ under H_1 has the cumulant generating function matching the cumulant generating function of $N(0, 1)$, neglecting the term $o(1/\sqrt{n})$. Put

$$c = -\frac{\mu\mu_3}{3\mu_2^2},$$

and consider the function

$$f(T^a) = \begin{cases} \frac{\mu}{c+1} \left(\left(\frac{T^a}{\mu} \right)^{c+1} - 1 \right), & (c \neq -1), \\ \mu \log \frac{T^a}{\mu}, & (c = -1). \end{cases}$$

The function f is not only strictly monotone for $T^a > 0$, and twice continuously differentiable at $T^a = \mu$, but it also satisfies the condition (3.3). Thus, for sufficiently large n , the distribution of $g(T^a)$ under H_1 is well approximated by $N(0, 1)$, that is,

$$(3.4) \quad \sqrt{\frac{n}{\mu_2}} \left\{ f(T^a) - \frac{1}{n} \left(\mu_1 - \frac{\mu_3}{6\mu_2} \right) \right\} \xrightarrow{L} N(0, 1), \quad \text{as } n \rightarrow \infty.$$

From (1.1), (2.1), and (3.4), we can approximate the power of $\mathbf{R}^a$ against H_1 as

$$(3.5) \quad 1 - \Phi \left(\sqrt{\frac{n}{\mu_2}} \left\{ f\left(\frac{t_0}{2n}\right) - \frac{1}{n} \left(\mu_1 - \frac{\mu_3}{6\mu_2} \right) \right\} \right),$$

where t_0 is the critical value of $\mathbf{R}^a$, and $\Phi(\cdot)$ denotes the standard normal distribution function.

The moment of $\mathbf{R}^a$ is infinite for $a \leq -1$. Therefore, the approximation given by (3.5) can not be defined for $a \leq -1$. However, since the probability that $g(T^a)$ is infinite goes to zero quickly, for sufficiently large n the distribution of $g(T^a)$ looks very much like $N(0, 1)$ (see Bishop et al. (1975, pp. 486-488)). For this reason, we propose to use the approximation given by (3.5) as well for $a \leq -1$.

4. Numerical comparisons of power approximations

In this section the normal power approximation based on a normalizing transformation proposed in Section 3 (NT approximation) is compared with the other approximations. In order to investigate the performance of the NT

approximation, the A^a approximation, the B^a approximation, the MCN approximation, and the ordinary noncentral χ^2 approximation (χ^2 approximation), we compare these approximations with the true power of the randomized size- α test that is calculated by direct enumeration (e.g. West and Kempthorne (1971) and Read and Cressie (1988, pp. 76-77)). The procedure for the comparison is as follows:

We consider $H_0: \mathbf{p}=(1/k, \dots, 1/k)'$ as the null hypothesis, and 500 alternatives $\mathbf{q}^{(i)}=(q_1^{(i)}, \dots, q_k^{(i)})'$ ($i=1, \dots, 500$) generated randomly. First, the true power of the randomized size- α test $PE(i)$ ($i=1, \dots, 500$) is calculated. Next, approximated power $P(i)$ ($i=1, \dots, 500$) is calculated by approximations of NT, A^a , B^a , MCN, and χ^2 respectively. As for the critical values, the corresponding critical values of the true power are used for all power approximations, to put all approximations on an equal footing. To calculate the approximated power by applying A^a , we need the upper probability of a linear combination of the noncentral χ^2 random variables. For this purpose, we use an algorithm developed by Sheil and O'Muircheartaigh (1977). The NT and MCN approximations can be correctly defined only for $\mathbf{R}^a(a > -1)$. However, we formally extend these approximations to $\mathbf{R}^a(a \leq -1)$. We compare the performance of the approximations based on the following two indices.

N_1 : The number of alternatives which satisfy $|P(i) - PE(i)| < 0.02$.

N_2 : The number of alternatives which satisfy $|P(i) - PE(i)| < 0.04$.

For sample size $n=20, 30, 40$, number of cells $k=4, 5$, and significance level $\alpha=0.05$, we consider statistics $\mathbf{R}^a$ with $a=-2, -1, -1/2, 0, 2/3, 1$. From Table 1 to Table 6, values of the indices N_1 and N_2 for each approximation are

Table 1. N_1 and N_2 for $k=4, n=20$

a	N_1					N_2				
	χ^2	MCN	A^a	B^a	NT	χ^2	MCN	A^a	B^a	NT
-2.0	0	42	48	34	*148	0	78	106	74	*259
-1.0	4	44	95	105	*294	14	69	205	225	*448
-0.5	44	54	*270	208	94	123	99	*369	337	262
0.0	158	108	179	179	*214	244	184	260	260	*347
2/3	190	172	*287	256	248	280	305	364	347	*383
1.0	97	66	241	*282	145	224	136	*489	448	245

Table 2. N_1 and N_2 for $k=4, n=30$

a	N_1					N_2				
	χ^2	MCN	A^a	B^a	NT	χ^2	MCN	A^a	B^a	NT
-2.0	2	34	65	52	*332	9	72	142	141	*497
-1.0	33	76	64	206	*291	47	139	171	335	*434
-0.5	68	63	344	361	*387	153	160	443	*490	461
0.0	164	186	238	238	*320	263	276	327	327	*439
2/3	205	168	*405	355	323	302	343	*492	453	437
1.0	220	137	*500	398	247	383	269	*500	491	379

Table 3. N_1 and N_2 for $k=4, n=40$

a	N_1					N_2				
	χ^2	MCN	A^a	B^a	NT	χ^2	MCN	A^a	B^a	NT
-2.0	27	44	90	126	*370	34	89	226	251	*485
-1.0	54	67	227	353	*400	132	133	415	*498	456
-0.5	88	103	322	*500	415	188	229	472	*500	457
0.0	137	164	327	327	*415	247	354	420	420	*455
2/3	241	191	*500	448	328	358	311	*500	485	416
1.0	232	149	*500	409	289	390	292	*500	481	377

Table 4. N_1 and N_2 for $k=5, n=20$

a	N_1					N_2				
	χ^2	MCN	A^a	B^a	NT	χ^2	MCN	A^a	B^a	NT
-0.5	0	28	141	*152	84	0	36	211	*225	178
0.0	0	17	117	117	*399	3	71	200	200	*477
2/3	33	244	*296	187	269	116	342	*498	356	365
1.0	153	176	*395	153	197	293	313	401	*440	348

Table 5. N_1 and N_2 for $k=5, n=30$

a	N_1					N_2				
	χ^2	MCN	A^a	B^a	NT	χ^2	MCN	A^a	B^a	NT
-2.0	0	*57	44	46	45	0	*111	78	105	94
-1.0	7	55	82	203	*335	11	111	177	378	*433
-0.5	19	83	172	298	*432	49	169	415	*497	492
0.0	50	164	195	195	*386	146	326	308	308	*467
2/3	204	179	*497	389	284	300	336	*500	495	411
1.0	217	125	*499	463	213	383	254	*500	497	355

Table 6. N_1 and N_2 for $k=5, n=40$

a	N_1					N_2				
	χ^2	MCN	A^a	B^a	NT	χ^2	MCN	A^a	B^a	NT
-2.0	3	24	59	93	*261	7	50	146	198	*476
-1.0	23	48	100	318	*344	36	117	269	438	*464
-0.5	39	82	207	375	*383	121	198	469	*500	436
0.0	132	196	258	258	*352	226	327	374	374	*420
2/3	234	203	*465	410	299	328	356	*500	487	397
1.0	179	158	*500	434	245	393	269	*500	498	326

listed. It is known that the critical value jumps to $+\infty$ as $\alpha \rightarrow 0$ for small n and $\alpha \leq -1$ (see Drost et al. (1989)). The critical values for $(k=5, n=20, \alpha=-2)$ and $(k=5, n=20, \alpha=-1)$ are too extreme. (Both of them are over 150.0 against the nominal value 9.49.) Then we omit the corresponding data from Table 4. The numerical value with the symbol * in the tables indicates the best value among five approximations for each $R^a(\alpha=-2, -1, -1/2, 0, 2/3, 1)$.

From these tables, we find the following results:

1. The improved normal approximation NT is superior to the MCN

approximation.

2. When $a=2/3, 1$, the A^a approximation is superior to the other approximations and is accurate when $n \geq 30$. The B^a approximation is the second best to A^a when $n \geq 30$.
3. When $a=0$, the NT approximation is superior to the other approximations in any of the cases stated above.
4. When $a < 0$, there are many cases where the NT approximation dominates the other approximations (especially in the cases of $a = -1, -2$).

5. Concluding remarks

From the results of numerical comparisons in Section 4, the NT approximation proposed in this paper is found to be superior to other approximations for the loglikelihood ratio statistic ($a=0$). It is also found that this approximation is very effective for the Freeman-Tukey statistic ($a=-1/2$), the modified loglikelihood ratio statistic ($a=-1$), and the Neyman-modified X^2 statistic ($a=-2$). Another merit of the NT approximation is that it is very easy to calculate. It needs only the upper probability of the standard normal distribution. On the other hand, the A^a approximation needs the eigenvalues of a certain matrix and the upper probability of a linear combination of noncentral χ^2 random variables. However, it is also true that the NT approximation is not performed well even in the case of $a=0$ compared to the A^a approximation performed when $a=1, 2/3$. The following are future subjects for inquiry.

1. Further studies of power approximation for the statistics $R^a (a \leq 0)$.
2. The reason why the NT approximation is not good when $a=1, 2/3$.
3. Comparison of the power of R^a using the approximations.

Acknowledgment

This research was partly supported by Grant-in-Aid for Scientific Research, No. 07558149, and No. 0940048, Ministry of Education, Science, Sports and Culture, Japan. The authors are very grateful to the referees and the editor for their valuable comments.

REFERENCES

- [1] Bishop, Y. M. M., Fienberg, S. E. and Holland, P. W. (1975). *Discrete multivariate analysis: Theory and practice*, The MIT Press.
- [2] Broffitt, J. D. and Randles, R. H. (1977). A power approximation for the chi-square goodness-of-fit test: simple hypothesis case, *J. Amer. Statist. Assoc.*, **72**, 604-607.
- [3] Cressie, N. and Read, T. R. C. (1984). Multinomial goodness-of-fit tests, *J. Roy. Statist. Soc., ser. B*, **46**, 440-464.
- [4] Drost, F. C., Kallenberg, W. C. M., Moore, D. S. and Oosterhoff, J. (1989). Power approximations to multinomial tests of fit, *J. Amer. Statist. Assoc.*, **84**, 130-141.
- [5] Konishi, S. (1981). Normalizing transformations of some statistics in multivariate analysis, *Biometrika*, **68**, 647-651.
- [6] Read, T. R. C. and Cressie, N. A. C. (1988). *Goodness-of-fit statistics for discrete multivariate data*, Springer.
- [7] Sekiya, Y. and Taneichi, N. (1994). On a normal approximation to the power of

- multinomial goodness-of-fit tests (in Japanese), *J. Hokkaido Univ. of Education* (2A), **45**, 1-4.
- [8] Sheil, J. and O'Muirheartaigh, I. (1977). As106. The distribution of non-negative quadratic forms in normal variables, *Appl. Statist.*, **26**, 92-98.
- [9] West, E. N. and Kempthorne, O. (1971). A comparison of χ^2 and likelihood ratio test for composite alternatives, *J. Statist. Comput. Simul.*, **1**, 1-33.