On performance of transformed ϕ-divergence goodness-of-fit statistic for testing logistic regression model

Nobuhiro Taneichi
Mathematics and Computer Science,Graduate School of Science and Engineering,
Kagoshima University, 1-21-35 Korimoto,
Kagoshima 890-0065,Japan
E-mail: taneichi@sci.kagoshima-u.ac.jp

Yuri Sekiya
Hokkaido University of Education,
1-15-55 Shiroyama,
Kushiro 085-8580, Japan
E-mail: sekiya.yuri@k.hokkyodai.ac.jp

Jun Toyama
Computer Science, Graduate School of Information Science and Technology,
Hokkaido University,
Kita 14 Nishi 9, Sapporo 060-0814, Japan
E-mail: jun@main.ist.hokudai.ac.jp

1. Introduction

We consider generalized linear models (Nelder and Wedderburn [4]) in which the response variables are measured on a binary scale. Let N independent random variables $Y_\alpha,\ \alpha = 1, \ldots, N$ corresponding to the number of successes in N different subgroups be distributed according to a binomial distribution $B(n_\alpha, \pi_\alpha),\ \alpha = 1, \ldots, N$. We consider the following general logistic regression model (general logit model):

$$\text{logit } \pi_\alpha = \boldsymbol{x}'_\alpha \boldsymbol{\beta}, \quad (\alpha = 1, \ldots, N), \tag{1}$$

where logit $u \equiv \log\{u/(1-u)\}$, $\boldsymbol{x}_\alpha = (x_{\alpha 1}, \ldots, x_{\alpha p})'$, $(\alpha = 1, \ldots, N)$, $(p < N)$ are covariate vectors and $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_p)'$ is a unknown parameter vector. In order to test null hypothesis

$$H_0 : \text{Model given by (1) is correct}, \tag{2}$$

we consider the family of ϕ-divergence statistics [7]

$$C_\phi = 2 \sum_{\alpha=1}^{N} n_\alpha \left\{ \hat{\pi}_\alpha \phi\left(\frac{Y_\alpha}{n_\alpha \hat{\pi}_\alpha}\right) + (1-\hat{\pi}_\alpha)\phi\left(\frac{n_\alpha - Y_\alpha}{n_\alpha(1-\hat{\pi}_\alpha)}\right) \right\},$$

where $\hat{\pi}_\alpha = \pi_\alpha(\hat{\boldsymbol{\beta}})$,$(\alpha = 1, \ldots, N)$, $\hat{\boldsymbol{\beta}} = (\hat{\beta}_1, \ldots, \hat{\beta}_p)'$ is the maximum likelihood estimator of $\boldsymbol{\beta}$ under H_0 given by (2) and $\phi(\cdot)$ is a real convex function in $(0, \infty)$, satisfying $\phi(1) = \phi'(1) = 0$ and $\phi''(1) = 1$. When we choose a convex function

$$\phi_a(t) = \begin{cases} \{a(a+1)\}^{-1}\{t^{a+1} - t + a(1-t)\} & (a \neq 0, -1) \\ t \log t + 1 - t & (a = 0) \\ -\log t - 1 + t & (a = -1) \end{cases} \tag{3}$$

as $\phi(t)$, C_{ϕ_a} becomes power divergence statistics [1]

$$R^a = 2 \sum_{\alpha=1}^{N} n_\alpha \left\{ I^a\left(\frac{Y_\alpha}{n_\alpha},\ \hat{\pi}_\alpha\right) + I^a\left(1 - \frac{Y_\alpha}{n_\alpha},\ 1 - \hat{\pi}_\alpha\right) \right\},$$

where,

$$I^a(e,f) = \begin{cases} \{a(a+1)\}^{-1} e\Big\{\Big(\frac{e}{f}\Big)^a - 1\Big\} & (a \neq 0, -1) \\ e \log\Big(\frac{e}{f}\Big) & (a = 0) \\ f \log\Big(\frac{f}{e}\Big) & (a = -1). \end{cases}$$

Under H_0, all members of the class of statistics C_ϕ have the χ^2_{N-p} limiting distribution, assuming a condition that

$$n_\alpha/n \to \mu_\alpha \ (0 < \mu_\alpha < 1) \text{ for each } \alpha, \quad \text{as} \quad n \to \infty, \tag{4}$$

where, $n = \sum_{\alpha=1}^N n_\alpha$ and $\sum_{\alpha=1}^N \mu_\alpha = 1$. Using large sample results, we can use C_ϕ for a goodness-of-fit test statistic for model (1). Statistic R^0 (log likelihood ratio statistic or deviance statistic) and statistic R^1 (Pearson's X^2 statistic) are used frequently.

However, in the case in which all n_α, $(\alpha = 1, \ldots, N)$ are not large enough, approximation by a χ^2_{N-p} limiting distribution to the distribution of C_ϕ under H_0 become poor. In such a case, there are risks that the hypothesis test based on large sample theory will give results opposite to those of an exact test. In order to reduce the risks, we propose a new transformed statistic $\tilde{C}_\phi$ of C_ϕ whose speed of convegence a chi-square distribution is quicker than C_ϕ.

2. Transformed ϕ-divergence statistics

First, we consider the following approximation based on an asymptotic expansion for the distribution of C_ϕ under H_0:

$$\Pr\{C_\phi \leq x | H_0\} \approx J_1^*(x) + J_2^*(x),$$

where the $J_1^*(x)$ term is multivariate Edgeworth expansion assuming a continuous distribution and the $J_2^*(x)$ term, which corresponds to the K_2 term of Taneichi et al. [5] in the case of a multinomial goodness-of-fit test, is a discontinuous term to account for the discontinuity. In this section, in order to evaluate $J_1^*(x)$ and $J_2^*(x)$, we need an assumption that the way of converging n_α/n to μ_α more strictly than the assumption given by (4). So, we consider the following Assumption 1 instead of the ssumption given by (4).

Assumption 1: $n_\alpha \to \infty$, $(\alpha = 1, \ldots, N)$, *as* $n \to \infty$, *with* n_α *depending on* n *in such a way that* $n_\alpha/n = \mu_\alpha$, $(\alpha = 1, \ldots, N)$, *where* $0 < \mu_\alpha < 1$ *and* $\sum_{\alpha=1}^N \mu_\alpha = 1$.

Under Assumption 1, the $J_1^*(x)$ term is evaluated as the following form.

$$J_1^*(x) = \Pr\{\chi^2_{N-p} \leq x\} + \frac{1}{n}\sum_{j=0}^{3} v_j^\phi \Pr\{\chi^2_{N-p+2j} \leq x\} + O(n^{-2}), \tag{5}$$

where χ^2_f denotes a chi-square random variable with degrees of freedom f. The coefficients v_j^ϕ, $(j = 0, \ldots, 3)$ do not depend on n and satisfy the relation $\sum_{j=0}^3 v_j^\phi = 0$. In the case of $\phi = \phi_0$, where ϕ_a is defined by (3), the coefficients $v_j^{\phi_0}$, $(j = 0, \ldots, 3)$ satisfy $v_1^{\phi_0} = -v_0^{\phi_0}$ and $v_2^{\phi_0} = v_3^{\phi_0} = 0$.

Though we omit the expression of $J_2^*(x)$, the $J_2^*(x)$ term is very complicated to calculate in practice. Therefore, we consider the approximation for the distribution of C_ϕ badsed on $J_1^*(x)$ term.

Next, using the term of multivariate Edgeworth expansion assuming a continuous distribution, we construct transformations for improving small-sample accuracy of the χ^2 approximation of the distribution of C_ϕ under H_0. The relation between coefficients of asymptotic expansion of a random variable and improved transformation of the random variable are shown as follows (e.g., Fijikoshi [2,3]). Suppose that a nonnegative random variable T has an asymptotic expansion such that

$$\Pr\{T \leq x\} = \Pr\{\chi^2_f \leq x\} + \frac{1}{n}\sum_{j=0}^{m} a_j \Pr\{\chi^2_{f+2j} \leq x\} + O(n^{-2}),$$

where m is a positive integer. The coefficients a_j, $(j = 0, \ldots, m)$ do not depend on the parameter $n > 0$ and must satisfy the relation $\sum_{j=0}^{m} a_j = 0$. For $m = 1$, we consider transformed random variable

$$T_B = \left(1 + \frac{2a_0}{fn}\right) T$$

which is known as the Bartlett adjustment. On the other hand, for $m = 3$, we consider transformed random variable

$$\begin{aligned} T_I &= (n\alpha + \beta)^2 \log \left[1 + \frac{1}{(n\alpha)^2} \left\{ T + \frac{1}{n\alpha}(T^2 + \gamma T^3) \right. \right. \\ &\qquad \left. \left. + \frac{1}{(n\alpha)^2} \left(\frac{1}{3}T^3 + \frac{3\gamma}{4}T^4 + \frac{9\gamma^2}{20}T^5 \right) \right\} \right], \end{aligned}$$

where $\alpha = -f(f+2)\{2(a_2 + a_3)\}^{-1}$, $\beta = -(f+2)a_0\{2(a_2 + a_3)\}^{-1}$ and $\gamma = a_3\{(f+4)(a_2 + a_3)\}^{-1}$. Then, it holds that

$$\Pr\{T_B \leq x\} = \Pr\{\chi_f^2 \leq x\} + O(n^{-2}) \quad \text{and} \quad \Pr\{T_I \leq x\} = \Pr\{\chi_f^2 \leq x\} + O(n^{-2}).$$

The proof of the results for improved transformation T_I is given by Yanagihara [6].

Applying the evaluation (5) to above transformed statistics T_B and T_I, we construct transformations for improving small-sample accuracy of the χ^2 approximation of the distribution of C_ϕ under H_0. When $\phi(\cdot)$ satisfies

$$\phi'''(1) = -1 \qquad \text{and} \qquad \phi^{(4)}(1) = 2, \tag{6}$$

we denote by $\tilde{C}_\phi^B$ the statistic based on Bartlett transformation for C_ϕ. On the other hand, when $\phi(\cdot)$ does not satisfy (6), we denote by $\tilde{C}_\phi^I$ the statistic based on improved transformation for C_ϕ. Then in the case of power divergence statistics $R^a = C_{\phi_a}$, we obtain transformed statistics $R_B^0 = \tilde{C}_{\phi_0}^B$ and $R_I^a = \tilde{C}_{\phi_a}^I, (a \neq 0)$.

3. Performance of transformed statistics

In order to investigate performance of the transformed statistics, we consider power divergence family of statistics as an example of ϕ-divergence family of statistics. We compare the performance of the transformed statistics R_B^0 and $R_I^a, (a \neq 0)$ and that of the original power divergence statistics R^a. We consider the logistic regression model given by (1) with $p = 2$ and $x_{\alpha 1} = 1$ and $x_{\alpha 2} = x_\alpha^*$, $(\alpha = 1, \ldots, N)$. Let the true values of parameters β_1 and β_2 be β_1^* and β_2^*, respectively. Then, the true value of π_α, $(\alpha = 1, \ldots, N)$ is

$$\pi_\alpha^* = \frac{\exp(\beta_1^* + \beta_2^* x_\alpha^*)}{1 + \exp(\beta_1^* + \beta_2^* x_\alpha^*)}, \qquad (\alpha = 1, \ldots, N).$$

We give a design matrix

$$\boldsymbol{X} = \begin{pmatrix} 1 & \cdots & 1 \\ x_1^* & \cdots & x_N^* \end{pmatrix}'$$

and execute the following procedure.

For each α, we generate n_α, $(\alpha = 1, \ldots, N)$ binomial random numbers which are distributed according to $B(1, \pi_\alpha^*)$, $(\alpha = 1, \ldots, N)$. From them, we calculate the number of success $Y_\alpha, (\alpha = 1, \ldots, N)$ and the maximum likelihood estimates $\hat{\beta}_1$ and $\hat{\beta}_2$ for the parameters β_1 and β_2. Using the estimates, we calculate the values $\pi_\alpha(\hat{\boldsymbol{\beta}})$, $(\alpha = 1, \ldots, N)$, where $\hat{\boldsymbol{\beta}} = (\hat{\beta}_1, \hat{\beta}_2)'$, and observed values of the statistics R^a, R_B^0 and $R_I^a, (a \neq 0)$. This process is repeated $J = 1.0 \times 10^6$ times. Among J times, let V be the number of times that the observed values of the statistics exceed the upper ε point of a chi-square distribution with degrees of freedom $N-p$, that is, $\chi^2_{N-p}(\varepsilon)$. The error of the χ^2 approximation for the distribution of each statistic can be evaluated on the basis of the index

$$I = \frac{V}{J} - \varepsilon.$$

We investigate the performance of the following two types when $N = 8$.

(I) True parameters are $\beta_1^* = 3, \beta_2^* = -8$, and a design matrix is

$$\boldsymbol{X} = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 0.2 & 0.25 & 0.3 & 0.35 & 0.4 & 0.45 & 0.5 & 0.55 \end{pmatrix}'.$$

(II) True parameters are $\beta_1^* = 4, \beta_2^* = -1$, and a design matrix is

$$\boldsymbol{X} = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 2.7 & 3.0 & 3.3 & 3.5 & 4.3 & 4.9 & 5.0 & 5.2 \end{pmatrix}'.$$

For each case, we consider sample design $n_1 = \cdots = n_8 = n_A$. Table 1 shows the values of $I \times 10^4$ for $a = 0$. From Table 1, we find that the performance of transformed statistic R_B^0 is better than that of original statistic R^0. For all cases, performance of transformed statistic R_B^0 is better than that of original statistic R^0. For almost all cases, the value of index I for R_B^0 is less than one-third of that for statistic R^0. As a result of comparison, we can say that statistic R^0 is improved by the transformed deviance statistic R_B^0.

Next, we consider the power of statistics R^a, R_B^0 and $R_I^a, (a \neq 0)$. We consider an alternative model:

$$\pi_\alpha^* = \frac{\exp(\beta_1^* + \beta_2^* x_\alpha^*)}{1 + \exp(\beta_1^* + \beta_2^* x_\alpha^*)} + \delta_\alpha, \qquad (\alpha = 1, \ldots, 8), \tag{7}$$

where $(\delta_1, \delta_2, \delta_3, \delta_4, \delta_5, \delta_6, \delta_7, \delta_8) = (-0.1,\ 0.1,\ -0.1,\ 0.1,\ -0.1,\ 0.1,\ -0.1,\ 0.1)$. We calculate the simulated power against the alternative model (7) by using simulated exact critical values of statistics R^a, R_B^0 and $R_I^a, (a \neq 0)$. Table 2 show the values of calculated power for $a = 0$. From Table 2, we find that the power of R_B^0 is not so different from the power of R^0.

4. Concluding remarks

We considered an approximation based on an asymptotic expansion for distribution of ϕ-divergence statistics C_ϕ in a logistic regression model. Using the continuous term of the expression of approximation for the distribution of ϕ-divergence statistics C_ϕ under a null hypothesis, we propose a transformation of C_ϕ that improves the speed of convergence to a chi-square limiting distribution. Numerical comparison shows that the transformed ϕ-divergence statistic is effective for improving the speed of convergence, especially, when $\phi = \phi_0$.

Table 1: Values of $I \times 10^4$ where $n_A = 5, 10, 15, 20, 30$ and significance level $\varepsilon = 0.01, 0.05, 0.1$ for cases (I)–(II) and $a = 0$.

		Type (I)							Type (II)				
$\varepsilon \setminus$	n_A	5	10	15	20	30	$\varepsilon \setminus$	n_A	5	10	15	20	30
0.01	R^0	126	87	57	42	20	0.01	R^0	118	93	61	41	20
	R^0_B	-50	-1	5	5	-1		R^0_B	-51	2	8	5	-2
0.05	R^0	513	310	196	133	67	0.05	R^0	524	338	206	132	68
	R^0_B	-136	21	26	15	-6		R^0_B	-138	42	36	15	-4
0.1	R^0	901	500	304	202	112	0.1	R^0	903	532	307	195	108
	R^0_B	-168	58	44	19	-4		R^0_B	-175	90	55	16	-7

Table 2: Power against alternative model (7) where $n_A = 5, 10, 15, 20, 30$ and significance level $\varepsilon = 0.01, 0.05, 0.1$ for cases (I)–(II) and $a = 0$.

		Type (I)							Type (II)				
$\varepsilon \setminus$	n_A	5	10	15	20	30	$\varepsilon \setminus$	n_A	5	10	15	20	30
0.01	R^0	0.04	0.08	0.16	0.24	0.46	0.01	R^0	0.04	0.09	0.16	0.25	0.48
	R^0_B	0.04	0.09	0.16	0.25	0.46		R^0_B	0.04	0.09	0.16	0.26	0.48
0.05	R^0	0.13	0.23	0.35	0.48	0.70	0.05	R^0	0.13	0.24	0.36	0.49	0.71
	R^0_B	0.13	0.24	0.36	0.48	0.70		R^0_B	0.14	0.24	0.36	0.49	0.71
0.1	R^0	0.21	0.35	0.48	0.61	0.80	0.1	R^0	0.22	0.35	0.49	0.62	0.81
	R^0_B	0.22	0.35	0.49	0.62	0.81		R^0_B	0.23	0.36	0.49	0.63	0.81

REFERENCES

[1] Cressie, N. and Read, T. R. C.: Multinomial goodness-of-fit tests, *J. R. Statist. Soc.* B, **46**, (1984), 440–464.

[2] Fujikoshi, Y.: A method for improving the large-sample chi-squared approximations to some multivariate test statistics. *Amer. J. Math. Management Sci.*, **17**, (1997) 15–29.

[3] Fujikoshi, Y.: Transformations with improved chi-squared approximations. *J. Mult. Anal.*, **72**, (2000), 249–263.

[4] Nelder, J. A. and Wedderburn, R. W. M.: Generalized linear models, *J. R. Statist. Soc.* A, **135**, (1972), 370–384.

[5] Taneichi, N., Sekiya, Y. and Suzukawa, A.: An asymptotic approximation for the distribution of $\phi-$divergence multinomial goodness-of-fit statistic under local alternatives, *J. Japan Statist. Soc.*, **31**(2), (2001), 207–224.

[5] Yanagihara, H.: Transformations for improving normal and chi-squared approximations, *Master's thesis, Hiroshima University*, (1998), (in Japanese).

[6] Zografos, K., Ferentions, K. and Papaioannou, T.: ϕ-divergence statistics: Sampling properties and multinomial goodness-of-fit and divergence tests, *Commun. Statist. -Theory and Meth.,* **19**, (1990), 1785–1802.

ABSTRACT

We consider logistic regression models in which the response variables are measured on a binary scale. We also consider a class of statistics which is based on ϕ-divergence as a goodness-of-fit test statistics for the model. The class of statistics include the statistics based on power divergence family as a special case. Well known Pearson's chi-square statistic and deviance (log likelihood ratio statistic) are included in power divergence family. All members of ϕ-divergence statistics have the limiting chi-square distribution assuming a certain condition under null hypothesis that logistic regression model is correct.

In this announcement, we construct and propose transformed ϕ-divergence statistics that improve the speed of convergence to the chi-square limiting distribution on the basis of the asymptotic expansion for lower probability of ϕ-divergence statistics. In order to investigate performance of the transformed statistic, we consider power divergence family of statistic as an example of ϕ-divergence family of statistic. Then, we numerically compare the speed of convergence of the transformed power divergence statistics with the original power divergence statistics. Furthermore, we also compare the power of the test based on transformed power divergence statistics with that of the test based on the original power divergence statistics.