

# Transformed goodness-of-fit statistics for a generalized linear model of binary data



Nobuhiro Taneichi<sup>a,\*</sup>, Yuri Sekiya<sup>b</sup>, Jun Toyama<sup>c</sup>

<sup>a</sup> Department of Mathematics and Computer Science, Graduate School of Science and Engineering, Kagoshima University, 1-21-35 Korimoto, Kagoshima 890-0065, Japan

<sup>b</sup> Kushiro Campus, Hokkaido University of Education, Kushiro 085-8580, Japan

<sup>c</sup> The Institute for Use of Mathematics, Sapporo 063-0001, Japan

## ARTICLE INFO

### Article history:

Received 13 October 2012

Available online 15 October 2013

### AMS 2000 subject classifications:

62E20

62H10

### Keywords:

Asymptotic expansion

Binary data

$\phi$ -divergence statistics

Generalized linear model

Improved transformation

## ABSTRACT

In a generalized linear model of binary data, we consider models based on a general link function including a logistic regression model and a probit model as special cases. For testing the null hypothesis  $H_0$  that the considered model is correct, we consider a family of  $\phi$ -divergence goodness-of-fit test statistics  $C_\phi$  that includes a power divergence family of statistics  $R^a$ . We propose a transformed  $C_\phi$  statistics that improves the speed of convergence to a chi-square limiting distribution and show numerically that the transformed  $R^a$  statistic performs well. We also give a real data example of the transformed  $R^a$  statistic being more reliable than the original  $R^a$  statistic for testing  $H_0$ .

© 2013 Elsevier Inc. All rights reserved.

## 1. Introduction

We consider generalized linear models [11] in which the response variables are measured on a binary scale. Let  $N$  independent random variables  $Y_\alpha$ ,  $\alpha = 1, \dots, N$  corresponding to the number of successes in  $N$  different subgroups be distributed according to binomial distributions  $B(n_\alpha, \pi_\alpha)$ ,  $\alpha = 1, \dots, N$ . If we use a monotone and differentiable function  $g(\cdot)$  as a link function, we obtain a generalized linear model for binary data as follows.

$$g(\pi_\alpha) = \mathbf{x}'_\alpha \boldsymbol{\beta}, \quad (\alpha = 1, \dots, N), \quad (1)$$

where  $\mathbf{x}_\alpha = (x_{\alpha 1}, \dots, x_{\alpha p})'$ ,  $(\alpha = 1, \dots, N)$ , are covariate vectors and  $\boldsymbol{\beta} = (\beta_1, \dots, \beta_p)'$  is an unknown parameter vector and  $p < N$ . When  $g(t)$  is a canonical link function, that is,

$$g(t) = \log \left( \frac{t}{1-t} \right), \quad (2)$$

model (1) is a logistic regression model. When

$$g(t) = g_p(t) = \Phi^{-1}(t), \quad (3)$$

\* Corresponding author.

E-mail addresses: [taneichi@sci.kagoshima-u.ac.jp](mailto:taneichi@sci.kagoshima-u.ac.jp) (N. Taneichi), [sekiya.yuri@k.hokkyodai.ac.jp](mailto:sekiya.yuri@k.hokkyodai.ac.jp) (Y. Sekiya), [mandheling@nifty.com](mailto:mandheling@nifty.com) (J. Toyama).

where  $\Phi(\cdot)$  is the cumulative distribution function of a standard normal distribution, model (1) is a probit model. When

$$g(t) = \log\{-\log(1-t)\}, \quad (4)$$

model (1) is a complementary log–log model. Aranda-Ordaz [1] considered a family of link functions,

$$g(t) = g_c(t) = \log \left\{ \frac{(1-t)^{-c} - 1}{c} \right\}, \quad (c \geq 0), \quad (5)$$

that depend on parameter  $c$ . By this family of link functions, we obtain a family of models that includes a logistic regression model when  $c = 1$  and a complementary log–log model when  $c = 0$  in the limit.

We consider the null hypothesis

$$H_0^g: \pi_\alpha = \pi_\alpha(\boldsymbol{\beta}) = g^{-1}(\boldsymbol{x}'_\alpha \boldsymbol{\beta}), \quad (\alpha = 1, \dots, N). \quad (6)$$

Here, we assume that nuisance parameters of (6) are only  $\boldsymbol{\beta}$ . That is, when we use  $g_c$  of (5) as a link function, we consider  $c$  to be fixed. In order to test the null hypothesis  $H_0^g$ , we consider the family of  $\phi$ -divergence statistics [13]

$$C_\phi = 2 \sum_{\alpha=1}^N n_\alpha \left\{ \hat{\pi}_\alpha^g \phi \left( \frac{\frac{Y_\alpha}{n_\alpha}}{\hat{\pi}_\alpha^g} \right) + (1 - \hat{\pi}_\alpha^g) \phi \left( \frac{1 - \frac{Y_\alpha}{n_\alpha}}{1 - \hat{\pi}_\alpha^g} \right) \right\}, \quad (7)$$

where  $\hat{\pi}_\alpha^g = \pi_\alpha(\hat{\boldsymbol{\beta}}^g)$ ,  $(\alpha = 1, \dots, N)$ ,  $\hat{\boldsymbol{\beta}}^g = (\hat{\beta}_1^g, \dots, \hat{\beta}_p^g)'$  is the maximum likelihood estimator of  $\boldsymbol{\beta}$  under  $H_0^g$  given by (6) and  $\phi(\cdot)$  is a real convex function in  $(0, \infty)$ , satisfying  $\phi(1) = \phi'(1) = 0$  and  $\phi''(1) = 1$ . Here, we note that test statistics  $C_\phi$  vary according to link function  $g$ . When we choose a convex function

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

as  $\phi(t)$ ,  $C_{\phi_a}$  becomes a power divergence statistic [5]

$$R^a = 2 \sum_{\alpha=1}^N n_\alpha \left\{ I^a \left( \frac{Y_\alpha}{n_\alpha}, \hat{\pi}_\alpha^g \right) + I^a \left( 1 - \frac{Y_\alpha}{n_\alpha}, 1 - \hat{\pi}_\alpha^g \right) \right\}, \quad (8)$$

where

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

Under  $H_0^g$ , all members of the class of statistics  $C_\phi$  have a  $\chi^2_{N-p}$  limiting distribution, assuming the condition that

$$n_\alpha/n \rightarrow \mu_\alpha, \quad (\alpha = 1, \dots, N) \text{ as } n \rightarrow \infty, \quad (9)$$

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

In the case of the goodness-of-fit test for a multinomial distribution, Yarnold [22] obtained an approximation based on asymptotic expansion for the null distribution of Pearson's  $\chi^2$  statistic. The expansion consists of a term of multivariate Edgeworth expansion for continuous distribution and discontinuous terms. Approximations based on asymptotic expansions for null distributions of some kinds of multinomial goodness-of-fit statistics have been investigated [16,14,10]. Edgeworth approximations of the distributions of some kinds of multinomial goodness-of-fit statistics under alternative hypotheses have also been investigated [18,17,15,12].

On the other hand, Taneichi et al. [19] obtained an approximation based on asymptotic expansion of the distribution of deviance for testing  $H_0^g$  given by (6) when link function  $g$  is defined by (2), that is, in a logistic regression model. Using the continuous term of the expression of the approximation, Taneichi et al. [19] proposed a Bartlett-type transformed statistic and showed that it improves the speed of convergence to a chi-square limiting distribution of the deviance.

In this paper, we first investigate asymptotic approximation of the distribution of the statistic  $C_\phi$  given by (7) for testing the null hypothesis  $H_0^g$  given by (6) in a general model given by (1). Next, for testing  $H_0^g$ , we propose a transformed  $C_\phi$  statistic that improves the speed of convergence to a chi-square limiting distribution. In Section 2, we first describe a local Edgeworth approximation for the probability of  $Y_\alpha$ , ( $\alpha = 1, \dots, N$ ) under  $H_0^g$ . Next, we consider expression of asymptotic expansion for the distribution of  $C_\phi$  under  $H_0^g$ . Evaluation for the continuous and discontinuous terms of the expression is considered. In Section 3, using the term of multivariate Edgeworth expansion assuming a continuous distribution in the expression in Section 2, we construct transformations for improving small-sample accuracy of the  $\chi^2$  approximation of the distribution of  $C_\phi$  under  $H_0^g$ . In Section 4, in the case of  $R^d$ , performance of the transformed statistic and that of the original statistic are compared numerically. In Section 5, we apply the transformed statistic to real data and discuss the importance of the transformed statistic.

## 2. Asymptotic approximation for the distribution of $C_\phi$ under $H_0^g$

First, we consider a local Edgeworth approximation for the probability of  $Y_\alpha$ , ( $\alpha = 1, \dots, N$ ) under null hypothesis  $H_0^g$  given by (6). Let  $Y_\alpha$ ,  $\alpha = 1, \dots, N$  be distributed according to a binomial distribution  $B(n_\alpha, \pi_\alpha^g)$ ,  $\alpha = 1, \dots, N$ , where each  $\pi_\alpha^g$ , ( $\alpha = 1, \dots, N$ ) is represented as  $\pi_\alpha^g = g^{-1}(\mathbf{x}'_\alpha \boldsymbol{\beta})$ , ( $\alpha = 1, \dots, N$ ) by using covariate vectors  $\mathbf{x}_\alpha = (x_{\alpha 1}, \dots, x_{\alpha p})'$ , and an unknown parameter vector  $\boldsymbol{\beta}$ . Let

$$W_\alpha = \frac{Y_\alpha - n_\alpha \pi_\alpha^g}{\sqrt{n_\alpha}}, \quad (\alpha = 1, \dots, N). \quad (10)$$

Then,  $\mathbf{W} = (W_1, \dots, W_N)'$  is a lattice random vector that takes values in the set

$$L = \left\{ \mathbf{w} = (w_1, \dots, w_N)' : w_\alpha = \frac{y_\alpha - n_\alpha \pi_\alpha^g}{\sqrt{n_\alpha}}, (\alpha = 1, \dots, N), \mathbf{y} = (y_1, \dots, y_N)' \in M \right\},$$

where

$$M = \left\{ \mathbf{y} = (y_1, \dots, y_N)' : y_1, \dots, y_N \text{ are non-negative integers that satisfy } y_\alpha \leq n_\alpha, (\alpha = 1, \dots, N) \right\}.$$

If we consider only for a limiting distribution of  $C_\phi$ , we can discuss under the assumption given by (9). In this section, since we consider asymptotic expansion of the distribution of  $C_\phi$ , we need an assumption that states the way of converging  $n_\alpha/n$  to  $\mu_\alpha$  more strictly than the assumption given by (9). Therefore, we consider the following **Assumption 1** instead of the assumption given by (9).

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

With regard to a local Edgeworth approximation for the probability of  $Y_\alpha$ , ( $\alpha = 1, \dots, N$ ) under  $H_0^g$ , we obtain the following lemma.

**Lemma 1.** For each  $\mathbf{y} = (y_1, \dots, y_N)' \in M$ , let  $\mathbf{w} = (w_1, \dots, w_N)'$ , where  $w_\alpha = (y_\alpha - n_\alpha \pi_\alpha^g)/\sqrt{n_\alpha}$ , ( $\alpha = 1, \dots, N$ ). Then, under **Assumption 1**,

$$\Pr\{\mathbf{W} = \mathbf{w} | H_0^g\} = \left( \prod_{\alpha=1}^N \frac{1}{\sqrt{n_\alpha}} \right) h^g(\mathbf{w}) \left\{ 1 + \frac{1}{\sqrt{n}} h_1^g(\mathbf{w}) + \frac{1}{n} h_2^g(\mathbf{w}) + \frac{1}{n\sqrt{n}} h_3^g(\mathbf{w}) + O(n^{-2}) \right\},$$

where

$$h^g(\mathbf{w}) = (2\pi)^{-N/2} |\Omega|^{-1/2} \exp\left(-\frac{1}{2} \mathbf{w}' \Omega^{-1} \mathbf{w}\right), \quad (11)$$

$$h_1^g(\mathbf{w}) = -\frac{1}{2} \sum_{\alpha=1}^N \frac{1}{\sqrt{\mu_\alpha}} \frac{(1-2\pi_\alpha^g)}{\pi_\alpha^g(1-\pi_\alpha^g)} w_\alpha + \frac{1}{6} \sum_{\alpha=1}^N \frac{1}{\sqrt{\mu_\alpha}} \frac{(1-2\pi_\alpha^g)}{(\pi_\alpha^g)^2(1-\pi_\alpha^g)^2} w_\alpha^3,$$

$$h_2^g(\mathbf{w}) = \frac{1}{2} \{h_1^g(\mathbf{w})\}^2 - \frac{1}{12} \sum_{\alpha=1}^N \frac{1}{\mu_\alpha} \frac{(1-\pi_\alpha^g + (\pi_\alpha^g)^2)}{\pi_\alpha^g(1-\pi_\alpha^g)} + \frac{1}{4} \sum_{\alpha=1}^N \frac{1}{\mu_\alpha} \frac{(1-2\pi_\alpha^g + 2(\pi_\alpha^g)^2)}{(\pi_\alpha^g)^2(1-\pi_\alpha^g)^2} w_\alpha^2 \\ - \frac{1}{12} \sum_{\alpha=1}^N \frac{1}{\mu_\alpha} \frac{(1-3\pi_\alpha^g + 3(\pi_\alpha^g)^2)}{(\pi_\alpha^g)^3(1-\pi_\alpha^g)^3} w_\alpha^4,$$

$$h_3^g(\mathbf{w}) = -\frac{1}{3} \{h_1^g(\mathbf{w})\}^3 + h_1^g(\mathbf{w}) h_2^g(\mathbf{w}) + \frac{1}{12} \sum_{\alpha=1}^N \frac{1}{\mu_\alpha \sqrt{\mu_\alpha}} \frac{(1-2\pi_\alpha^g)}{(\pi_\alpha^g)^2(1-\pi_\alpha^g)^2} w_\alpha \\ - \frac{1}{6} \sum_{\alpha=1}^N \frac{1}{\mu_\alpha \sqrt{\mu_\alpha}} \frac{(1-2\pi_\alpha^g)(1-\pi_\alpha^g + (\pi_\alpha^g)^2)}{(\pi_\alpha^g)^3(1-\pi_\alpha^g)^3} w_\alpha^3 + \frac{1}{20} \sum_{\alpha=1}^N \frac{1}{\mu_\alpha \sqrt{\mu_\alpha}} \frac{(1-2\pi_\alpha^g)(1-2\pi_\alpha^g + 2(\pi_\alpha^g)^2)}{(\pi_\alpha^g)^4(1-\pi_\alpha^g)^4} w_\alpha^5,$$

and

$$\Omega = \text{diag}(\pi_1^g(1 - \pi_1^g), \dots, \pi_N^g(1 - \pi_N^g)). \quad (12)$$

The proof of Lemma 1 is similar to that for Lemma 1 of [19], which is proved by considering the proof of Theorem 22.1 of [4, pp. 232–236].

Next, we derive an approximation based on an asymptotic expansion for the distribution of  $C_\phi$  under  $H_0^g$ . We consider the following approximation for the distribution of  $C_\phi$  under  $H_0^g$  corresponding to approximation (2.3) of [17] for the goodness-of-fit test.

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

where the  $J_1^{g,\phi}(x)$  term is multivariate Edgeworth expansion assuming a continuous distribution and the  $J_2^{g,\phi}(x)$  term, which corresponds to the  $K_2$  term of [17] in the case of a multinomial goodness-of-fit test, is a discontinuous term to account for the discontinuity. With regard to evaluation of the  $J_1^{g,\phi}(x)$  term, we obtain the following theorem.

**Theorem 1.** When  $g^{-1}$  is a fourth time continuously differentiable function and  $\phi$  is a fifth time continuously differentiable function, under Assumption 1, the  $J_1^{g,\phi}(x)$  term is evaluated as

$$J_1^{g,\phi}(x) = \Pr\{\chi_{N-p}^2 \leq x\} + \frac{1}{n} \sum_{j=0}^3 v_j^{g,\phi} \Pr\{\chi_{N-p+2j}^2 \leq x\} + O(n^{-2}), \quad (13)$$

where  $\chi_f^2$  denotes a chi-square random variable with degrees of freedom  $f$ ,

$$\begin{aligned} v_0^{g,\phi} &= \frac{1}{24}(-\Gamma_4), \\ v_1^{g,\phi} &= \frac{1}{24}[\Gamma_1\phi^{(4)}(1) + \Gamma_2\{\phi'''(1) + 1\}^2 + (2\Gamma_1 + \Gamma_3)\phi'''(1) + (\Gamma_3 + \Gamma_4)], \\ v_2^{g,\phi} &= \frac{1}{24}[-\Gamma_1\phi^{(4)}(1) - 2\Gamma_2\{\phi'''(1) + 1\}^2 - (2\Gamma_1 + \Gamma_3)\phi'''(1) - \Gamma_3], \\ v_3^{g,\phi} &= \frac{1}{24}\Gamma_2\{\phi'''(1) + 1\}^2, \end{aligned}$$

where

$$\begin{aligned} \Gamma_1 &= -3(A_1 - 2A_3 + A_6), \quad \Gamma_2 = 5A_2 - 12A_4 + 9A_7 - 3B_1 + 6B_2 - 2B_4 - 3B_7, \\ \Gamma_3 &= 2(3A_1 - 2A_2 - 6A_3 + 6A_4 + 3A_5 + 3A_6 - 6A_7 - 3A_8 - 3B_3 + 2B_4 + 3B_8), \\ \Gamma_4 &= 6A_1 - 4A_2 - 6A_6 + 12A_8 - 3A_9 + 4B_4 - 12B_5 + 6B_6 - 3B_9, \\ A_1 &= \sum_{\alpha=1}^N \frac{1 - 3\pi_\alpha^g + 3(\pi_\alpha^g)^2}{\mu_\alpha \pi_\alpha^g (1 - \pi_\alpha^g)}, \quad A_2 = \sum_{\alpha=1}^N \frac{(1 - 2\pi_\alpha^g)^2}{\mu_\alpha \pi_\alpha^g (1 - \pi_\alpha^g)}, \\ A_3 &= \sum_{\alpha=1}^N \frac{1 - 3\pi_\alpha^g + 3(\pi_\alpha^g)^2}{(\pi_\alpha^g)^2 (1 - \pi_\alpha^g)^2} G_1(\alpha)^2 \sigma_{\alpha\alpha}, \quad A_4 = \sum_{\alpha=1}^N \frac{(1 - 2\pi_\alpha^g)^2}{(\pi_\alpha^g)^2 (1 - \pi_\alpha^g)^2} G_1(\alpha)^2 \sigma_{\alpha\alpha}, \\ A_5 &= \sum_{\alpha=1}^N \frac{(1 - 2\pi_\alpha^g)}{\pi_\alpha^g (1 - \pi_\alpha^g)} G_2(\alpha) \sigma_{\alpha\alpha}, \quad A_6 = \sum_{\alpha=1}^N \frac{\mu_\alpha (1 - 3\pi_\alpha^g + 3(\pi_\alpha^g)^2)}{(\pi_\alpha^g)^3 (1 - \pi_\alpha^g)^3} G_1(\alpha)^4 \sigma_{\alpha\alpha}^2, \\ A_7 &= \sum_{\alpha=1}^N \frac{\mu_\alpha (1 - 2\pi_\alpha^g)^2}{(\pi_\alpha^g)^3 (1 - \pi_\alpha^g)^3} G_1(\alpha)^4 \sigma_{\alpha\alpha}^2, \quad A_8 = \sum_{\alpha=1}^N \frac{\mu_\alpha (1 - 2\pi_\alpha^g)}{(\pi_\alpha^g)^2 (1 - \pi_\alpha^g)^2} G_1(\alpha)^2 G_2(\alpha) \sigma_{\alpha\alpha}^2, \\ A_9 &= \sum_{\alpha=1}^N \frac{\mu_\alpha}{\pi_\alpha^g (1 - \pi_\alpha^g)} G_2(\alpha)^2 \sigma_{\alpha\alpha}^2, \\ B_1 &= \sum_{\alpha=1}^N \sum_{\gamma=1}^N \frac{(1 - 2\pi_\alpha^g)}{\pi_\alpha^g (1 - \pi_\alpha^g)} \frac{(1 - 2\pi_\gamma^g)}{\pi_\gamma^g (1 - \pi_\gamma^g)} G_1(\alpha) G_1(\gamma) \sigma_{\alpha\gamma}, \\ B_2 &= \sum_{\alpha=1}^N \sum_{\gamma=1}^N \frac{\mu_\alpha (1 - 2\pi_\alpha^g)}{(\pi_\alpha^g)^2 (1 - \pi_\alpha^g)^2} \frac{(1 - 2\pi_\gamma^g)}{\pi_\gamma^g (1 - \pi_\gamma^g)} G_1(\alpha)^3 G_1(\gamma) \sigma_{\alpha\alpha} \sigma_{\alpha\gamma}, \end{aligned}$$

$$\begin{aligned}
B_3 &= \sum_{\alpha=1}^N \sum_{\gamma=1}^N \frac{\mu_{\alpha}}{\pi_{\alpha}^g(1-\pi_{\alpha}^g)} \frac{(1-2\pi_{\gamma}^g)}{\pi_{\gamma}^g(1-\pi_{\gamma}^g)} G_1(\alpha)G_2(\alpha)G_1(\gamma)\sigma_{\alpha\alpha}\sigma_{\alpha\gamma}, \\
B_4 &= \sum_{\alpha=1}^N \sum_{\gamma=1}^N \frac{\mu_{\alpha}(1-2\pi_{\alpha}^g)}{(\pi_{\alpha}^g)^2(1-\pi_{\alpha}^g)^2} \frac{\mu_{\gamma}(1-2\pi_{\gamma}^g)}{(\pi_{\gamma}^g)^2(1-\pi_{\gamma}^g)^2} G_1(\alpha)^3 G_1(\gamma)^3 \sigma_{\alpha\gamma}^3, \\
B_5 &= \sum_{\alpha=1}^N \sum_{\gamma=1}^N \frac{\mu_{\alpha}}{\pi_{\alpha}^g(1-\pi_{\alpha}^g)} \frac{\mu_{\gamma}(1-2\pi_{\gamma}^g)}{(\pi_{\gamma}^g)^2(1-\pi_{\gamma}^g)^2} G_1(\alpha)G_2(\alpha)G_1(\gamma)^3 \sigma_{\alpha\gamma}^3, \\
B_6 &= \sum_{\alpha=1}^N \sum_{\gamma=1}^N \frac{\mu_{\alpha}}{\pi_{\alpha}^g(1-\pi_{\alpha}^g)} \frac{\mu_{\gamma}}{\pi_{\gamma}^g(1-\pi_{\gamma}^g)} G_1(\alpha)G_2(\alpha)G_1(\gamma)G_2(\gamma)\sigma_{\alpha\gamma}^3, \\
B_7 &= \sum_{\alpha=1}^N \sum_{\gamma=1}^N \frac{\mu_{\alpha}(1-2\pi_{\alpha}^g)}{(\pi_{\alpha}^g)^2(1-\pi_{\alpha}^g)^2} \frac{\mu_{\gamma}(1-2\pi_{\gamma}^g)}{(\pi_{\gamma}^g)^2(1-\pi_{\gamma}^g)^2} G_1(\alpha)^3 G_1(\gamma)^3 \sigma_{\alpha\alpha}\sigma_{\alpha\gamma}\sigma_{\gamma\gamma}, \\
B_8 &= \sum_{\alpha=1}^N \sum_{\gamma=1}^N \frac{\mu_{\alpha}}{\pi_{\alpha}^g(1-\pi_{\alpha}^g)} \frac{\mu_{\gamma}(1-2\pi_{\gamma}^g)}{(\pi_{\gamma}^g)^2(1-\pi_{\gamma}^g)^2} G_1(\alpha)G_2(\alpha)G_1(\gamma)^3 \sigma_{\alpha\alpha}\sigma_{\alpha\gamma}\sigma_{\gamma\gamma}, \\
B_9 &= \sum_{\alpha=1}^N \sum_{\gamma=1}^N \frac{\mu_{\alpha}}{\pi_{\alpha}^g(1-\pi_{\alpha}^g)} \frac{\mu_{\gamma}}{\pi_{\gamma}^g(1-\pi_{\gamma}^g)} G_1(\alpha)G_2(\alpha)G_1(\gamma)G_2(\gamma)\sigma_{\alpha\alpha}\sigma_{\alpha\gamma}\sigma_{\gamma\gamma}, \\
G_i(\alpha) &= u^{(i)}(\mathbf{x}'_{\alpha}\boldsymbol{\beta}), \quad (\alpha = 1, \dots, N, i = 1, 2), \\
u(x) &= g^{-1}(x), \\
\sigma_{\alpha\gamma} &= \sum_{l=1}^p \sum_{m=1}^p \kappa^{l,m} x_{\alpha l} x_{\gamma m}, \quad (\alpha, \gamma = 1, \dots, N), \\
\kappa_{l,m} &= \sum_{\lambda=1}^N \mu_{\lambda} \{\pi_{\lambda}^g(1-\pi_{\lambda}^g)\}^{-1} G_1(\lambda)^2 x_{\lambda l} x_{\lambda m}, \quad (l, m = 1, \dots, p),
\end{aligned}$$

where  $u^{(i)}$  is the  $i$ th derivative of  $u$  and  $\kappa^{l,m}$  is the  $(l, m)$ -element of the inverse matrix  $K^{-1}$  of  $K = (\kappa_{l,m})$ .

Proof of [Theorem 1](#) is shown in [Appendix A](#). We find that the coefficients  $v_j^{g,\phi}$ , ( $j = 0, 1, 2, 3$ ) satisfy the relation  $\sum_{j=0}^3 v_j^{g,\phi} = 0$ . If we apply  $\phi_a$  as  $\phi$  in [Theorem 1](#), we obtain the following corollary for the statistic  $R^a$  based on power divergence.

**Corollary 1.** When the statistic is  $R^a$  given by (8) and  $g^{-1}$  is a fourth time continuously differentiable function, under [Assumption 1](#), the  $J_1^{g,\phi}(x)$  term is evaluated as

$$J_1^{g,\phi}(x) = \Pr\{\chi_{N-p}^2 \leq x\} + \frac{1}{n} \sum_{j=0}^3 v_j^{g,(a)} \Pr\{\chi_{N-p+2j}^2 \leq x\} + O(n^{-2}), \quad (14)$$

where  $v_j^{g,(a)}$ , ( $j = 0, 1, 2, 3$ ) are defined as  $v_j^{g,\phi}$ , ( $j = 0, 1, 2, 3$ ) in the case of  $\phi'''(1) = a - 1$  and  $\phi^{(4)}(1) = (a - 1)(a - 2)$ , respectively.

We give some examples with regard to [Theorem 1](#). When link function  $g(\cdot)$  is given by (2), that is, the logistic regression model, then

$$\begin{aligned}
g^{-1}(x) &= \frac{\exp(x)}{1 + \exp(x)}, \\
G_1(\alpha) &= \pi_{\alpha}^g(1 - \pi_{\alpha}^g), \quad (\alpha = 1, \dots, N),
\end{aligned}$$

and

$$G_2(\alpha) = \pi_{\alpha}^g(1 - \pi_{\alpha}^g)(1 - 2\pi_{\alpha}^g), \quad (\alpha = 1, \dots, N)$$

in [Theorem 1](#). Applying the above results to [Corollary 1](#) in the case of  $a = 0$ , the expansion (14) coincides with expansion (2.5) of [19], that is, an expansion for the deviance for the logistic regression model. When link function  $g(\cdot)$  is given by (3),

that is, the probit model, then

$$g^{-1}(x) = \Phi(x),$$

$$G_1(\alpha) = \frac{1}{\sqrt{2\pi}} \exp \left[ -\frac{\{\Phi^{-1}(\pi_\alpha^g)\}^2}{2} \right], \quad (\alpha = 1, \dots, N),$$

and

$$G_2(\alpha) = -\frac{1}{\sqrt{2\pi}} \Phi^{-1}(\pi_\alpha^g) \exp \left[ -\frac{\{\Phi^{-1}(\pi_\alpha^g)\}^2}{2} \right], \quad (\alpha = 1, \dots, N)$$

in Theorem 1. When link function  $g(\cdot)$  is given by (4), that is, the complementary log–log model, then

$$g^{-1}(x) = 1 - \exp\{-\exp(x)\},$$

$$G_1(\alpha) = -(1 - \pi_\alpha^g) \log(1 - \pi_\alpha^g), \quad (\alpha = 1, \dots, N),$$

and

$$G_2(\alpha) = -(1 - \pi_\alpha^g) \{\log(1 - \pi_\alpha^g)\} \{1 + \log(1 - \pi_\alpha^g)\}, \quad (\alpha = 1, \dots, N)$$

in Theorem 1. When link function  $g(\cdot)$  is given by (5), that is, the family of models proposed by [1], then

$$g^{-1}(x) = 1 - \{1 + c \exp(x)\}^{-\frac{1}{c}}, \quad (c \geq 0),$$

$$G_1(\alpha) = c^{-1} (1 - \pi_\alpha^g) \{1 - (1 - \pi_\alpha^g)^c\}, \quad (c \geq 0, \alpha = 1, \dots, N),$$

and

$$G_2(\alpha) = c^{-2} (1 - \pi_\alpha^g) \{1 - (1 - \pi_\alpha^g)^c\} \{(c+1)(1 - \pi_\alpha^g)^c - 1\}, \quad (c \geq 0, \alpha = 1, \dots, N)$$

in Theorem 1.

Next, we consider the  $J_2^{g,\phi}(x)$  term. Let  $U_\phi^g(x)$  be a set defined by

$$U_\phi^g(x) = \{\mathbf{w} = (w_1, \dots, w_N)' : C_\phi(\mathbf{w}) \leq x\}. \quad (15)$$

Define the sets  $U_\gamma^{g,\phi} (\subset R_{N-1})$ ,  $(\gamma = 1, \dots, N)$  and continuous functions  $\eta_\gamma^{g,\phi}(\cdot)$  and  $\theta_\gamma^{g,\phi}(\cdot)$ ,  $(\gamma = 1, \dots, N)$  on  $R_{N-1}$  into  $R_1$  such that  $U_\phi^g(x)$  defined by (15) is represented as

$$U_\phi^g(x) = \{\mathbf{w} = (w_1, \dots, w_N)' : \eta_\gamma^{g,\phi}(\tilde{\mathbf{w}}_\gamma) \leq w_\gamma \leq \theta_\gamma^{g,\phi}(\tilde{\mathbf{w}}_\gamma), \tilde{\mathbf{w}}_\gamma = (w_1, \dots, w_{\gamma-1}, w_{\gamma+1}, \dots, w_N)' \in U_\gamma^{g,\phi}\}, \quad (16)$$

where  $R_M$  is an  $M$ -dimensional Euclidean space. Then  $J_2^{g,\phi}(x)$  is defined using the sets  $U_\gamma^{g,\phi}$ ,  $(\gamma = 1, \dots, N)$  given by (16) as follows.

$$J_2^{g,\phi}(x) = -\frac{1}{\sqrt{n}} \sum_{\gamma=1}^N n^{-(N-\gamma)/2} \sum_{w_{\gamma+1} \in L_{\gamma+1}} \cdots \sum_{w_N \in L_N} \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \chi_{U_\gamma^{g,\phi}}(\tilde{\mathbf{w}}_\gamma) \\ \times [S_1(\sqrt{n}w_\gamma + n\pi_\gamma^g) h^g(\mathbf{w})]_{\eta_\gamma^{g,\phi}(\tilde{\mathbf{w}}_\gamma)}^{\theta_\gamma^{g,\phi}(\tilde{\mathbf{w}}_\gamma)} dw_1 \cdots dw_{\gamma-1}, \quad (17)$$

where

$$[F(\mathbf{w})]_{\eta_\gamma^{g,\phi}(\tilde{\mathbf{w}}_\gamma)}^{\theta_\gamma^{g,\phi}(\tilde{\mathbf{w}}_\gamma)} = F(w_1, \dots, w_{\gamma-1}, \theta_\gamma^{g,\phi}(\tilde{\mathbf{w}}_\gamma), w_{\gamma+1}, \dots, w_N) - F(w_1, \dots, w_{\gamma-1}, \eta_\gamma^{g,\phi}(\tilde{\mathbf{w}}_\gamma), w_{\gamma+1}, \dots, w_N),$$

$$L_\gamma = \left\{ w_\gamma : w_\gamma = \frac{y_\gamma - n_\gamma \pi_\gamma^g}{\sqrt{n_\gamma}}, y_\gamma \text{ is a non-negative integer that satisfies } y_\gamma \leq n_\gamma \right\}, \quad (\gamma = 1, \dots, N), \quad (18)$$

$$S_1(t) = t - [t] - \frac{1}{2}, \quad (19)$$

$h^g(\cdot)$  being defined by (11), and  $\chi_A(\cdot)$  being the indicate function of the set  $A$ . In order to evaluate the  $J_2^{g,\phi}(x)$  term of the null distribution of the test statistic using the same method as that of [22], it is necessary to show

$$[S_1(\sqrt{n}w_\gamma + n\pi_\gamma^g) h^g(\mathbf{w})]_{\eta_\gamma^{g,\phi}(\tilde{\mathbf{w}}_\gamma)}^{\theta_\gamma^{g,\phi}(\tilde{\mathbf{w}}_\gamma)} = b [S_1(\sqrt{n}w_\gamma + n\pi_\gamma^g)]_{\eta_\gamma^{g,\phi}(\tilde{\mathbf{w}}_\gamma)}^{\theta_\gamma^{g,\phi}(\tilde{\mathbf{w}}_\gamma)} + o(1),$$

where  $b$  is a constant. However, it is very difficult to show the above relation except when  $h^g(\mathbf{w})$  is a constant. Therefore, unlike the null distribution of multinomial goodness-of-fit test statistics, we cannot obtain a simple form of approximation of the  $J_2^{g,\phi}(x)$  term such as  $\hat{K}_2$  given by (2.6) of [17]. By another method of [22],  $J_2^{g,\phi}(x)$  is evaluated as follows.

**Theorem 2.** When  $g^{-1}$  is a fourth time continuously differentiable function and  $\phi$  is a fifth time continuously differentiable function, under [Assumption 1](#), the  $J_2^{g,\phi}(x)$  term can be represented in the following form:

$$J_2^{g,\phi}(x) = \left\{ (2\pi)^N \prod_{\alpha=1}^N \pi_{\alpha}^g (1 - \pi_{\alpha}^g) \right\}^{-1/2} (\Theta_1^{g,\phi} + \Theta_2^{g,\phi} - \Theta_3^{g,\phi} + O(n^{-2})), \quad (20)$$

where

$$\begin{aligned} \Theta_1^{g,\phi} &= n^{-N/2} \sum_{w_1 \in L_1} \cdots \sum_{\substack{w_N \in L_N \\ \mathbf{w} \in U_{\phi}^g(x)}} \exp\left(-\frac{1}{2} \mathbf{w}' \Omega^{-1} \mathbf{w}\right), \\ \Theta_2^{g,\phi} &= \frac{1}{\sqrt{n}} \frac{1}{\pi_N^g (1 - \pi_N^g)} \int \cdots \int_{U_{\phi}^g(x)} w_N S_1(\sqrt{n} w_N + n \pi_N^g) \exp\left(-\frac{1}{2} \mathbf{w}' \Omega^{-1} \mathbf{w}\right) d\mathbf{w} + \frac{1}{n} \frac{1}{\pi_{N-1}^g (1 - \pi_{N-1}^g)} \\ &\quad \times \sum_{w_N \in L_N} \int \cdots \int_{Q_N(w_N)} w_{N-1} S_1(\sqrt{n} w_{N-1} + n \pi_{N-1}^g) \exp\left(-\frac{1}{2} \mathbf{w}' \Omega^{-1} \mathbf{w}\right) dw_1 \cdots dw_{N-1} \\ &\quad + \frac{1}{n\sqrt{n}} \frac{1}{\pi_{N-2}^g (1 - \pi_{N-2}^g)} \sum_{w_{N-1} \in L_{N-1}} \sum_{w_N \in L_N} \int \cdots \int_{Q_{N-1,N}(w_{N-1}, w_N)} 1 \\ &\quad \times S_1(\sqrt{n} w_{N-2} + n \pi_{N-2}^g) \exp\left(-\frac{1}{2} \mathbf{w}' \Omega^{-1} \mathbf{w}\right) dw_1 \cdots dw_{N-2}, \end{aligned}$$

$$\Theta_3^{g,\phi} = \Pr\{\chi_{N-p}^2 \leq x\} + \frac{1}{n} \sum_{j=0}^3 \zeta_j^{g,\phi} \Pr\{\chi_{N-p+2j}^2 \leq x\},$$

$$Q_N(w_N) = \{(w_1, \dots, w_{N-1})' : \mathbf{w} = (w_1, \dots, w_{N-1}, w_N)' \in U_{\phi}^g(x)\},$$

$$Q_{N-1,N}(w_{N-1}, w_N) = \{(w_1, \dots, w_{N-2})' : \mathbf{w} = (w_1, \dots, w_{N-2}, w_{N-1}, w_N)' \in U_{\phi}^g(x)\},$$

$$\zeta_0^{g,\phi} = \frac{1}{24}(-\Gamma_7), \quad \zeta_1^{g,\phi} = \frac{1}{24}\{\Gamma_1 \phi^{(4)} + \Gamma_2 (\phi'''(1))^2 + \Gamma_5 \phi'''(1) + (\Gamma_6 + \Gamma_7)\},$$

$$\zeta_2^{g,\phi} = \frac{1}{24}\{-\Gamma_1 \phi^{(4)}(1) - 2\Gamma_2 (\phi'''(1))^2 - \Gamma_5 \phi'''(1) - \Gamma_6\}, \quad \zeta_3^{g,\phi} = \frac{1}{24} \Gamma_2 (\phi'''(1))^2,$$

$$\Gamma_5 = 6(A_5 - A_8 - B_3 + B_8), \quad \Gamma_6 = -3(2A_4 - 4A_7 + B_1 - 2B_2 + 2B_4 + B_7),$$

$$\Gamma_7 = -3(-4A_3 + 2A_4 + 2A_5 + 4A_6 - 10A_8 + A_9 + B_1 - 2B_2 - 2B_3 - 2B_4 + 8B_5 - 2B_6 + B_7 + 2B_8 + B_9),$$

where  $d\mathbf{w} = dw_1 \cdots dw_N$ ,  $A_3, \dots, A_9$ ,  $B_1, \dots, B_9$  and  $\Gamma_1$  and  $\Gamma_2$  are given in [Theorem 1](#).

The proof of [Theorem 2](#) is shown in [Appendix B](#). By [Theorem 2](#), we find that the  $J_2^{g,\phi}(x)$  term is very difficult to calculate in practice. Therefore, we consider approximation for the distribution of  $C_{\phi}$  based on the  $J_1^{g,\phi}(x)$  term.

### 3. Transformed statistic based on the $J_1^{g,\phi}(x)$ term

In this section, we first describe the idea of transformation for improving small-sample accuracy of  $\chi^2$  approximations of the distribution of a random variable.

Suppose that a nonnegative random variable  $T$  has an asymptotic expansion such that

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

where  $m$  is a positive integer. Also suppose that the coefficients  $a_j$ , ( $j = 0, 1, \dots, m$ ) do not depend on the parameter  $n(>0)$  and must satisfy the relation  $\sum_{j=0}^m a_j = 0$ .

For  $m = 1$ , in order to increase the accuracy of  $\chi^2$  approximation of a random variable  $T$ , we consider transformed random variable  $T_B$  defined by

$$T_B = \left(1 + \frac{2a_0}{fn}\right) T. \quad (21)$$

Then, it holds that

$$\Pr\{T_B \leq x\} = \Pr\{\chi_f^2 \leq x\} + O(n^{-2}). \quad (22)$$

This result is known as a Bartlett adjustment. Lawley [9], Barndorff-Nielsen and Cox [2], and Barndorff-Nielsen and Hall [3] discussed Bartlett adjustment for the log likelihood ratio statistic.

For  $m = 3$ , in order to increase the accuracy of  $\chi^2$  approximation of a random variable  $T$ , we consider transformed random variable  $T_l$  defined by

$$T_l = (n\alpha + \beta)^2 \log \left[ 1 + \frac{1}{(n\alpha)^2} \left\{ T + \frac{1}{n\alpha} (T^2 + \gamma T^3) + \frac{1}{(n\alpha)^2} \left( \frac{1}{3} T^3 + \frac{3\gamma}{4} T^4 + \frac{9\gamma^2}{20} T^5 \right) \right\} \right], \quad (23)$$

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

$$\Pr\{T_l \leq x\} = \Pr\{\chi_f^2 \leq x\} + O(n^{-2}).$$

The proof of the results for transformation of  $T_l$  is given by [21]. The proof is derived by applying the idea of [8] to the theory of improved transformation given by [7].

Applying the evaluation (13) given by Theorem 1 to the above transformed statistics  $T_B$  given by (21) and  $T_l$  given by (23), we construct transformations for improving small-sample accuracy of the  $\chi^2$  approximation of the distribution of  $C_\phi$  under  $H_0^g$ .

When  $\phi(\cdot)$  satisfies

$$\phi'''(1) = -1 \quad \text{and} \quad \phi^{(4)}(1) = 2, \quad (24)$$

equations  $v_1^{g,\phi} = -v_0^{g,\phi}$  and  $v_2^{g,\phi} = v_3^{g,\phi} = 0$  hold in Theorem 1. Then, we can consider Bartlett-type adjustment

$$C_\phi^{B*} = \left\{ 1 + \frac{2v_0^{g,\phi}}{n(N-p)} \right\} C_\phi.$$

On the other hand, when  $\phi(\cdot)$  does not satisfy (24), we can consider transformed statistic

$$C_\phi^{I*} = (n\alpha + \beta)^2 \log(1 + \zeta),$$

where

$$\zeta = \frac{1}{(n\alpha)^2} \left[ C_\phi + \frac{1}{n\alpha} \{(C_\phi)^2 + \gamma(C_\phi)^3\} + \frac{1}{(n\alpha)^2} \left\{ \frac{1}{3}(C_\phi)^3 + \frac{3\gamma}{4}(C_\phi)^4 + \frac{9\gamma^2}{20}(C_\phi)^5 \right\} \right],$$

$\alpha = -(N-p)(N-p+2)\{2(v_2^{g,\phi} + v_3^{g,\phi})\}^{-1}$ ,  $\beta = -(N-p+2)v_0^{g,\phi}\{2(v_2^{g,\phi} + v_3^{g,\phi})\}^{-1}$  and  $\gamma = v_3^{g,\phi}\{(N-p+4)(v_2^{g,\phi} + v_3^{g,\phi})\}^{-1}$ .

Practically, we may use estimate  $\hat{v}_j^{g,\phi}$ , ( $j = 0, 2, 3$ ) obtained by substituting maximum likelihood estimate  $\hat{\beta}^g$  for true value  $\beta$  in  $v_j^{g,\phi}$ , ( $j = 0, 2, 3$ ). Therefore, when  $\phi(\cdot)$  satisfies (24), we propose the statistic  $\tilde{C}_\phi^B$  which is obtained by substituting  $\hat{v}_0^{g,\phi}$  for  $v_0^{g,\phi}$  in  $C_\phi^{B*}$ , that is,

$$\tilde{C}_\phi^B = \left\{ 1 + \frac{2\hat{v}_0^{g,\phi}}{n(N-p)} \right\} C_\phi. \quad (25)$$

Similarly, when  $\phi(\cdot)$  does not satisfy (24), we also propose the statistic  $\tilde{C}_\phi^I$  which is obtained by substituting  $\hat{v}_j^{g,\phi}$ , ( $j = 0, 2, 3$ ) for  $v_j^{g,\phi}$ , ( $j = 0, 2, 3$ ) in  $C_\phi^{I*}$ .

In the case of power divergence statistic  $R^a = C_{\phi_a}$ , condition (24) is satisfied if and only if  $a = 0$  (log likelihood ratio statistic). Then, we consider the transformed statistic given by (25) when  $a = 0$  and put  $\tilde{R}_B^0 = \tilde{C}_{\phi_0}^B$ . When the link function  $g(\cdot)$  is defined by (2), that is, logistic regression model, statistic  $\tilde{R}_B^0$  coincides with the statistic  $\tilde{D}$  proposed by (3.4) of [19]. On the other hand, we consider statistic  $\tilde{C}_{\phi_a}^I$  when  $a \neq 0$  and put  $\tilde{R}_I^a = \tilde{C}_{\phi_a}^I$ , ( $a \neq 0$ ).

In [2], the theory of Bartlett adjustment is discussed for the case in which the error term in (22) is not  $O(n^{-2})$  but  $O(n^{-3/2})$ . In Theorem 1, we evaluated the  $J_1^{g,\phi}(x)$  term up to order  $n^{-3/2}$ . Therefore, we can apply the continuous part of asymptotic expansion for the distribution of  $R^0$  (log likelihood ratio statistic) to the above formula when  $m = 1$ , which ensures better accuracy of approximation than the theory of [2].

#### 4. Performance of transformed statistics

In this section, we compare the performance of transformed statistics  $\tilde{R}_I^a$ , ( $a \neq 0$ ) and  $\tilde{R}_B^0$  with that of the original power divergence statistics  $R^a$  using the Monte Carlo procedure. We consider a generalized linear model given by (1) with  $p = 2$  and  $x_{\alpha,1} = 1$  and  $x_{\alpha,2} = x_\alpha$ , ( $\alpha = 1, \dots, N$ ).



Let the true values of parameters  $\beta_1$  and  $\beta_2$  be  $\beta_1^*$  and  $\beta_2^*$ , respectively. Then, the true value of  $\pi_\alpha^g$ , ( $\alpha = 1, \dots, N$ ) is

$$\pi_\alpha^{g*} = g^{-1}(\beta_1^* + \beta_2^* x_\alpha), \quad (\alpha = 1, \dots, N). \quad (26)$$

As a link function  $g(\cdot)$ , we consider the family of link functions  $g_c(\cdot)$  ( $0 \leq c \leq 1$ ) given by (5) including the logit link  $g_1(\cdot)$  and complementary log–log link  $g_0(\cdot)$ . We also consider the probit link  $g_p(\cdot)$  given by (3).

We give a design matrix

$$\mathbf{X} = \begin{pmatrix} 1 & \cdots & 1 \\ x_1 & \cdots & x_N \end{pmatrix}'$$

and execute the following procedure.

For each  $\alpha$ , we generate  $n_\alpha$ , ( $\alpha = 1, \dots, N$ ) binomial random numbers that are distributed according to  $B(1, \pi_\alpha^{g*})$ , ( $\alpha = 1, \dots, N$ ). From them, we calculate the number of successes  $Y_\alpha$ , ( $\alpha = 1, \dots, N$ ) and the maximum likelihood estimates  $\hat{\beta}_1^g$  and  $\hat{\beta}_2^g$  for the parameters  $\beta_1$  and  $\beta_2$  by Fisher scoring. Using the estimates, we calculate the values  $\pi_\alpha(\hat{\beta}^g)$ , ( $\alpha = 1, \dots, N$ ), where  $\hat{\beta}^g = (\hat{\beta}_1^g, \hat{\beta}_2^g)'$ , and observed values of the statistics  $R^a$ ,  $\tilde{R}_l^a$  ( $a \neq 0$ ), and  $\tilde{R}_b^0$ . This process is repeated  $D$  times.

Among  $D$  times, let  $V$  be the number of times that the observed values of the statistic exceed the upper  $\varepsilon$  point of the  $\chi^2$  distribution with degrees of freedom  $N - p$ , that is,  $\chi_{N-p}^2(\varepsilon)$ . The performance of  $\chi^2$  approximation for the distribution of each statistic can be evaluated on the basis of the index

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

We consider the following two true parameters

- (i)  $\beta_1^* = -0.1, \beta_2^* = 0.1$ ,
- (ii)  $\beta_1^* = 0.1, \beta_2^* = -0.1$ ,

and investigate the performance of the following four cases of design matrix when  $N = 8$ .

(I)

$$\mathbf{X} = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 2.7 & 3.0 & 3.3 & 3.6 & 3.9 & 4.2 & 4.5 & 4.8 \end{pmatrix}'.$$

(II)

$$\mathbf{X} = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 2.85 & 3.05 & 3.25 & 3.45 & 3.65 & 3.85 & 4.05 & 4.25 \end{pmatrix}'.$$

(III)

$$\mathbf{X} = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ \log(2.7) & \log(3.0) & \log(3.3) & \log(3.6) & \log(3.9) & \log(4.2) & \log(4.5) & \log(4.8) \end{pmatrix}'.$$

(IV)

$$\mathbf{X} = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ \log(2.85) & \log(3.05) & \log(3.25) & \log(3.45) & \log(3.65) & \log(3.85) & \log(4.05) & \log(4.25) \end{pmatrix}'.$$

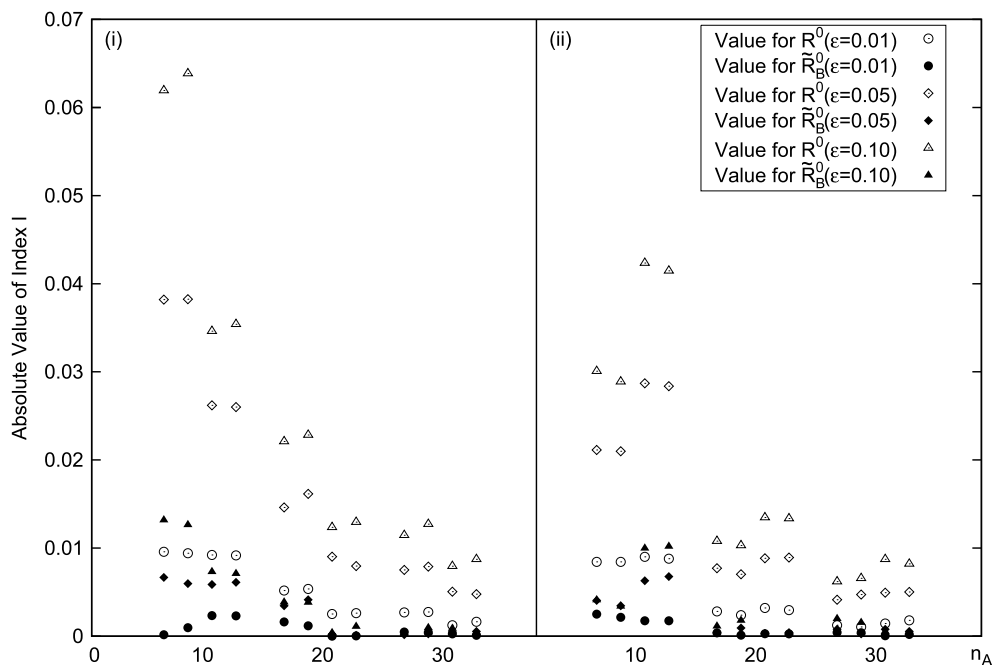
For each case, we consider the following two sample designs.

(A)  $n_1 = \dots = n_8 = n_A$ .

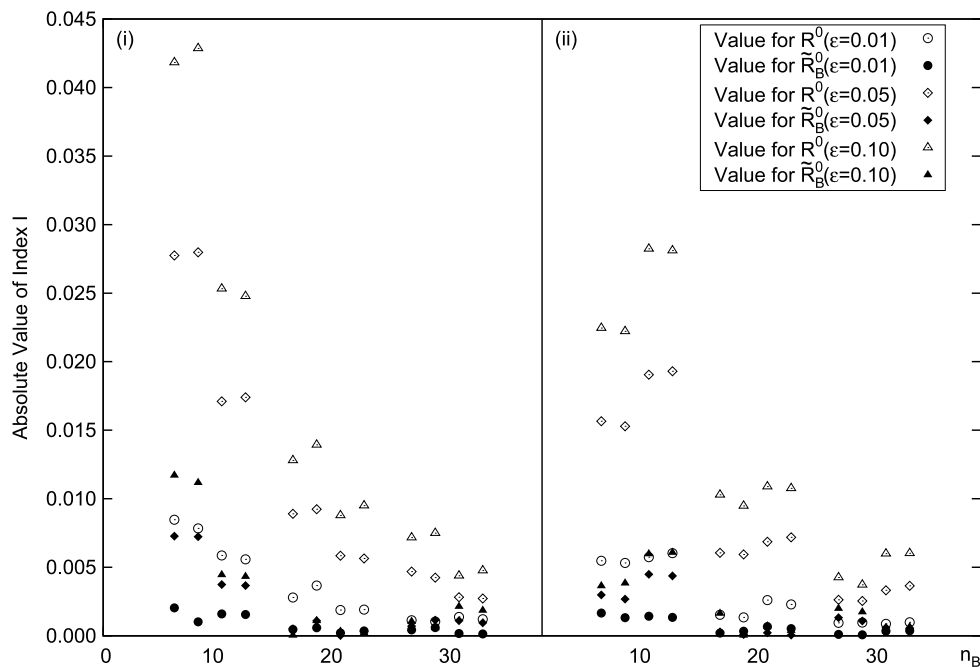
(B)  $n_1 = \dots = n_4 = n_B, n_5 = \dots = n_8 = 2n_B$ .

We investigate the performance for all combinations of two true parameters (i) and (ii), four design matrices (I), (II), (III), and (IV), and sample design (A), where  $n_A = 10, 20$ , and  $30$ , and sample design (B), where  $n_B = 10, 20$ , and  $30$ . In the investigation, the number of repetitions is  $D = 10^6$ . Parts of the results of the investigations are shown in figures as follows.

Figs. 1 and 2 show the absolute values of index  $I$  in the cases of true parameters (i) and (ii), design matrices (I)–(IV) and significance level  $\varepsilon = 0.01, 0.05$ , and  $0.10$  when the model is given by the link function  $g_0(\cdot)$  (complementary log–log model) and the original test statistic is log likelihood ratio statistic  $R^0$ . Fig. 1 is the case for sample design (A), where  $n_A = 10, 20$ , and  $30$  and Fig. 2 is the case for sample design (B), where  $n_B = 10, 20$ , and  $30$ . Fig. 3 shows the absolute values of index  $I$  when the model is given by probit link function  $g_p(\cdot)$  in the same situation as that in the explanation of Fig. 1. Fig. 4 shows the absolute values of index  $I$  when the test statistic is  $R^{0.2}$  and the model is given by link function  $g_1(\cdot)$  (logistic regression model),  $g_0(\cdot)$  (complementary log–log model),  $g_{1/2}(\cdot)$ , and  $g_p(\cdot)$  (probit model) in the case of true parameters (i) and (ii), design matrices (I)–(IV), and sample design (A), where  $n_A = 10$ . Fig. 5 shows the absolute values of index  $I$  when the test statistic is  $R^{2/3}$  in the same models and situations as those in the explanation of Fig. 4.

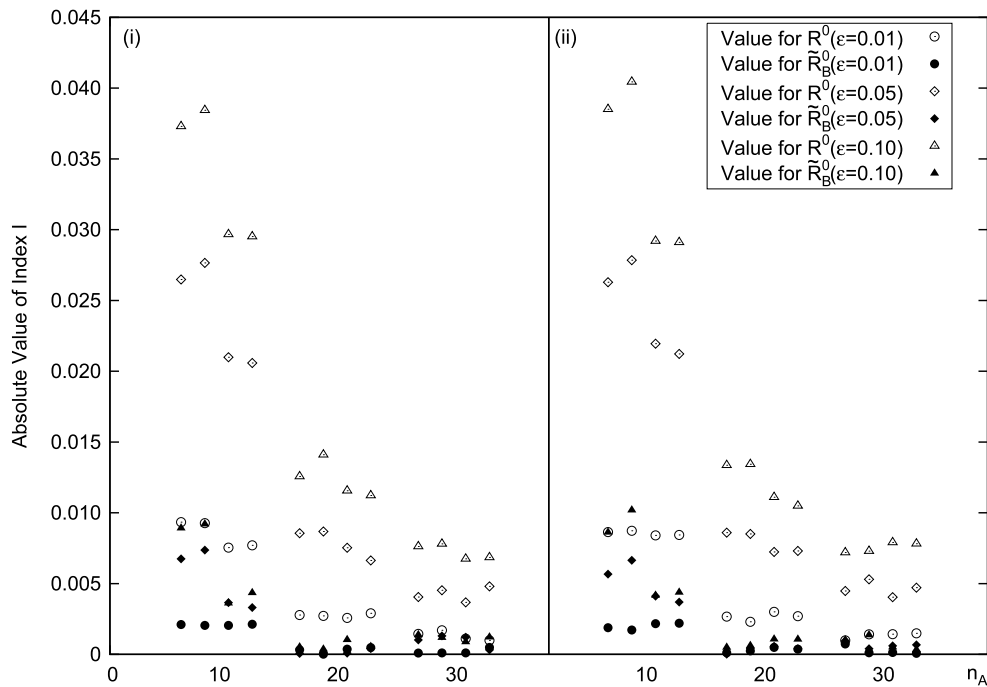


**Fig. 1.** Absolute value of index  $I$  when the model is given by complementary log–log link function  $g_0(\cdot)$  and original statistic is  $R^0$  (log likelihood ratio statistic) for true parameters (i) and (ii) and sample design (A) with  $n_A = 10, 20, 30$ :  $\circ, \diamond$  and  $\triangle$  are the values for  $R^0$  when  $\varepsilon = 0.01, 0.05$  and  $0.10$ , respectively, and  $\bullet, \blacklozenge$  and  $\blacktriangle$  are the values for  $\tilde{R}_B^0$  when  $\varepsilon = 0.01, 0.05$  and  $0.10$ , respectively. The 1st column is for design matrix (I), the 2nd column is for design matrix (II), the 3rd column is for design matrix (III), and the 4th column is for design matrix (IV).

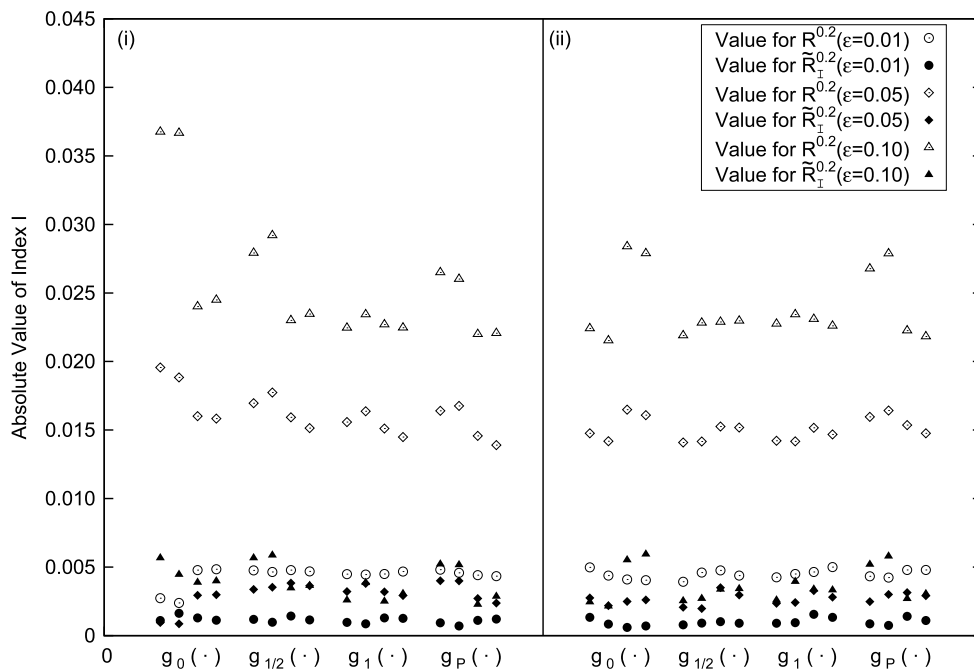


**Fig. 2.** Absolute value of index  $I$  when the model is given by complementary log–log link function  $g_0(\cdot)$  and original statistic is  $R^0$  (log likelihood ratio statistic) for true parameters (i) and (ii) and sample design (B) with  $n_B = 10, 20, 30$ :  $\circ, \diamond$  and  $\triangle$  are the values for  $R^0$  when  $\varepsilon = 0.01, 0.05$  and  $0.10$ , respectively, and  $\bullet, \blacklozenge$  and  $\blacktriangle$  are the values for  $\tilde{R}_B^0$  when  $\varepsilon = 0.01, 0.05$  and  $0.10$ , respectively. The 1st column is for design matrix (I), the 2nd column is for design matrix (II), the 3rd column is for design matrix (III), and the 4th column is for design matrix (IV).

From Figs. 1 and 2, we find that the performance of transformed statistic  $\tilde{R}_B^0$  is better than that of original statistic  $R^0$  in a model given by link function  $g_0(\cdot)$  (complementary log–log model) for two true parameters, all design matrix cases, and both sample designs (A) and (B). From Fig. 3, we find that the performance of transformed statistic  $\tilde{R}_B^0$  is better than that of original

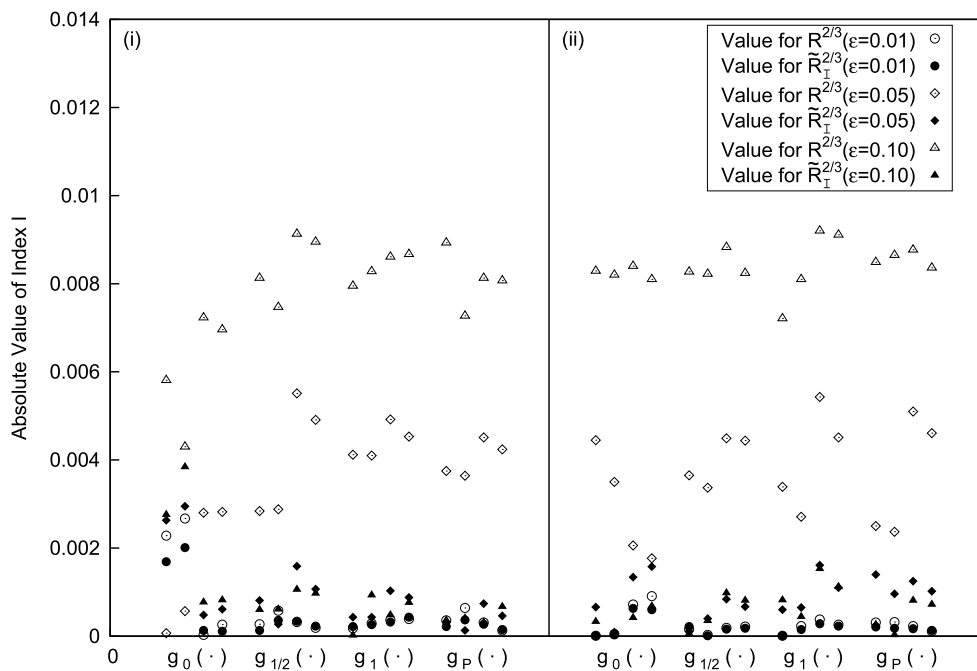


**Fig. 3.** Absolute value of  $I$  when the model is given by probit link function  $g_P(\cdot)$  and original statistic is  $R^0$  (log likelihood ratio statistic) for true parameters (i) and (ii) and sample design (A) with  $n_A = 10, 20, 30$ :  $\circ, \diamond$  and  $\triangle$  are the values for  $R^0$  when  $\varepsilon = 0.01, 0.05$  and  $0.10$ , respectively, and  $\bullet, \blacklozenge$  and  $\blacktriangle$  are the values for  $\tilde{R}_B^0$  when  $\varepsilon = 0.01, 0.05$  and  $0.10$ , respectively. The 1st column is for design matrix (I), the 2nd column is for design matrix (II), the 3rd column is for design matrix (III), and the 4th column is for design matrix (IV).



**Fig. 4.** Absolute value of index  $I$  when original test statistic is  $R^{0.2}$  and the model is given by link function  $g_0(\cdot), g_{1/2}(\cdot), g_1(\cdot)$  and  $g_P(\cdot)$  for true parameters (i) and (ii) and sample design (A) with  $n_A = 10, 20, 30$ :  $\circ, \diamond$  and  $\triangle$  are the values for  $R^{0.2}$  when  $\varepsilon = 0.01, 0.05$  and  $0.10$ , respectively, and  $\bullet, \blacklozenge$  and  $\blacktriangle$  are the values for  $\tilde{R}_I^{0.2}$  when  $\varepsilon = 0.01, 0.05$  and  $0.10$ , respectively. The 1st column is for design matrix (I), the 2nd column is for design matrix (II), the 3rd column is for design matrix (III), and the 4th column is for design matrix (IV).

statistic  $R^0$  in the model given by link function  $g_P(\cdot)$  (probit model) for two true parameters, all design matrix cases, and sample design (A). In the case of sample design (B), provided that the other situations and the model are the same as those of



**Fig. 5.** Absolute value of index  $I$  when original test statistic is  $R^{2/3}$  and the model is given by link function  $g_0(\cdot)$ ,  $g_{1/2}(\cdot)$ ,  $g_1(\cdot)$  and  $g_P(\cdot)$  for true parameters (i) and (ii) and sample design (A) with  $n_A = 10$ :  $\circ$ ,  $\diamond$  and  $\triangle$  are the values for  $R^{2/3}$  when  $\varepsilon = 0.01, 0.05$  and  $0.10$ , respectively, and  $\bullet$ ,  $\blacklozenge$  and  $\blacktriangle$  are the values for  $\tilde{R}_I^{2/3}$  when  $\varepsilon = 0.01, 0.05$  and  $0.10$ , respectively. The 1st column is for design matrix (I), the 2nd column is for design matrix (II), the 3rd column is for design matrix (III), and the 4th column is for design matrix (IV).

Fig. 3, the results of investigations are similar to those of Fig. 3. Simulation results of comparison of the performance between  $\tilde{R}_B^0$  and  $R^0$  in the model given by link function  $g_1(\cdot)$  (logistic regression model) are shown in [19], and it was concluded that the performance of  $\tilde{R}_B^0$  is also better than that of  $R^0$ . From Fig. 4, we find that the performance of transformed statistic  $\tilde{R}_I^{0.2}$  is better than that of original statistic  $R^{0.2}$  for two true parameters, all design matrix cases, and sample design (A), where  $n_A = 10$ . In the case of sample design (A), where  $n_A = 20$  and  $30$ , and sample design (B), where  $n_B = 10, 20$  and  $30$ , provided that the other situations and the models are the same as those of Fig. 4, the results of investigations are similar to those of Fig. 4. From Fig. 5, we find that the performance of transformed statistic  $\tilde{R}_I^{2/3}$  is better than that of original statistic  $R^{2/3}$  when the models are given by the link functions  $g_{1/2}(\cdot)$ ,  $g_1(\cdot)$  (logistic regression model) and  $g_P(\cdot)$  (probit model) for two true parameters, all design matrix cases, and sample design (A), where  $n_A = 10$ . However, when the model is given by the link function  $g_0(\cdot)$  (complementary log–log model), the performance of transformed statistic  $\tilde{R}_I^{2/3}$  is not better than that of original statistic  $R^{2/3}$ . Similarly, under the condition of the same situations and models as those of Figs. 4 and 5, we investigated the performance of  $\tilde{R}_I^1$  and  $R^1$ . From the results of the investigation, we found that the transformed statistic performs better than the original statistic in logistic regression and probit models. However, the transformed statistic does not perform better than the original statistic in the complementary log–log model.

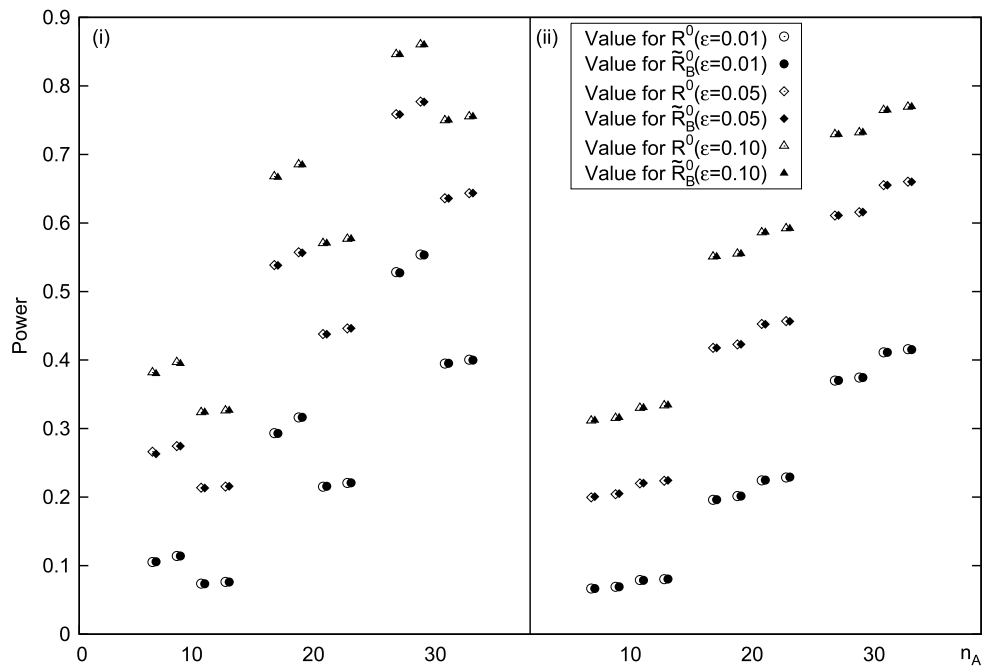
Consequently, from Figs. 1–5 and the other simulation results, we conclude as follows. Performance of  $\tilde{R}_B^0$  is better than that of original statistic  $R^0$ . When the statistic is  $R^{0.2}$ , performance of  $\tilde{R}_I^{0.2}$  is also better than that of original statistic  $R^{0.2}$ . When the statistic is  $R^a$  ( $a = 2/3, 1$ ), the transformed statistic performs better than the original statistic in models including the logistic regression model and the probit model. However, when the statistic is also  $R^a$  ( $a = 2/3, 1$ ), in the complementary log–log model, the transformed statistic does not perform better than the original statistic in some situations.

Next, we compare the power of transformed statistics  $\tilde{R}_I^a$ , ( $a \neq 0$ ) and  $\tilde{R}_B^0$  with that of the original statistics  $R^a$ . Against the null model given by (26), we consider an alternative model:

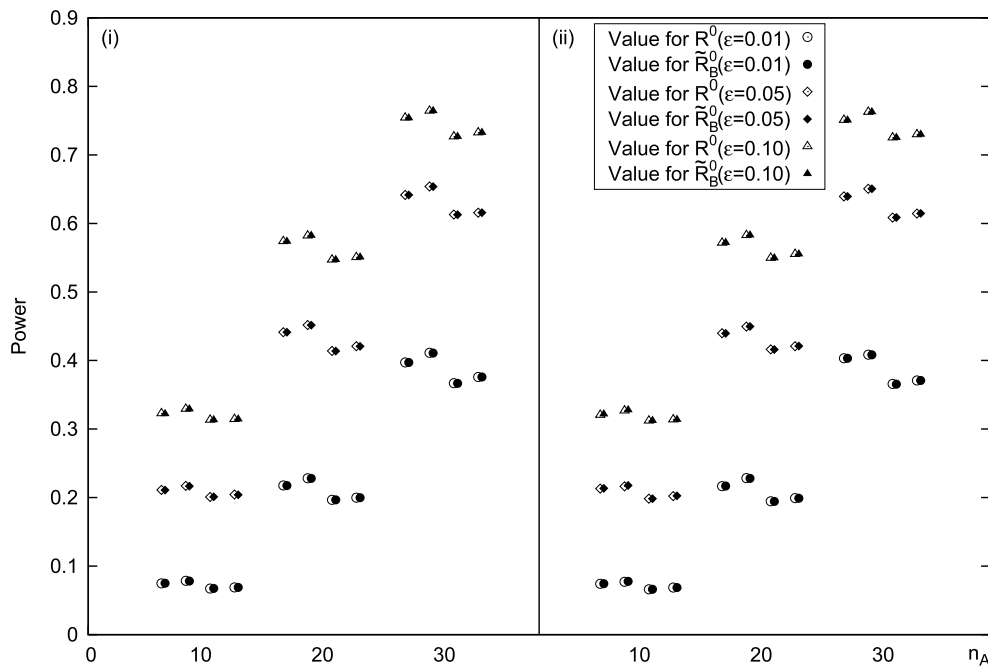
$$H_1^g: \pi_{\alpha}^{g*} = g^{-1}(\beta_1^* + \beta_2^* x_{\alpha}) + \delta_{\alpha}, \quad (\alpha = 1, \dots, 8), \quad (27)$$

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 average power  $P$  against the alternative model (27) by using simulated exact critical values of statistics. We investigate the average power for all combinations of two true parameters (i) and (ii), four design matrices (I)–(IV), and sample design (A), where  $n_A = 10, 20$ , and  $30$  and sample design (B), where  $n_B = 10, 20$ , and  $30$ . In the investigation, the number of repetitions is  $D = 10^6$ . Parts of the results of the investigations are shown in figures as follows. Figs. 6–8 show the power of statistics corresponding to the cases of Figs. 1, 3 and 5, respectively.

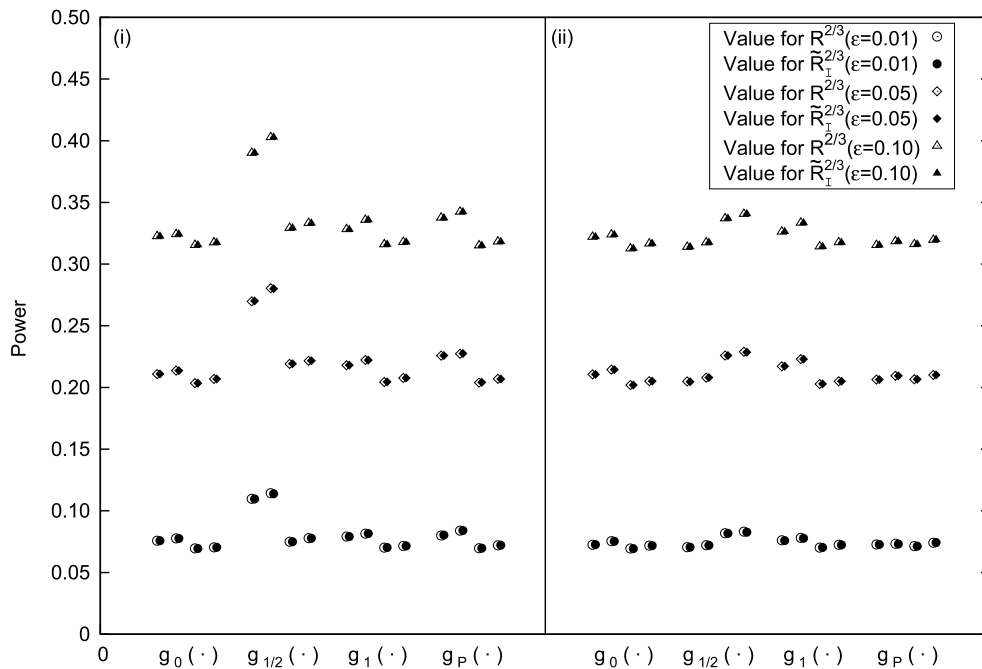


**Fig. 6.** Simulated average power  $P$  against an alternative model (27) when the model is given by complementary log–log link function  $g_0(\cdot)$  and original statistic is  $R^0$  (log likelihood ratio statistic) for true parameters (i) and (ii) and sample design (A) with  $n_A = 10, 20, 30$ :  $\circ, \diamond$  and  $\triangle$  are the values for  $R^0$  when  $\varepsilon = 0.01, 0.05$  and  $0.10$ , respectively, and  $\bullet, \blacklozenge$  and  $\blacktriangle$  are the values for  $R_B^0$  when  $\varepsilon = 0.01, 0.05$  and  $0.10$ , respectively. The 1st column is for design matrix (I), the 2nd column is for design matrix (II), the 3rd column is for design matrix (III), and the 4th column is for design matrix (IV).



**Fig. 7.** Simulated average power  $P$  against an alternative model (27) when the model is given by probit link function  $g_P(\cdot)$  and original statistic is  $R^0$  (log likelihood ratio statistic) for true parameters (i) and (ii) and sample design (A) with  $n_A = 10, 20, 30$ :  $\circ, \diamond$  and  $\triangle$  are the values for  $R^0$  when  $\varepsilon = 0.01, 0.05$  and  $0.10$ , respectively, and  $\bullet, \blacklozenge$  and  $\blacktriangle$  are the values for  $R_B^0$  when  $\varepsilon = 0.01, 0.05$  and  $0.10$ , respectively. The 1st column is for design matrix (II), the 3rd column is for design matrix (III), and the 4th column is for design matrix (IV).

From Figs. 6–8, and the other simulation results, we conclude that the power against  $H_1^\varepsilon$  given by (27) of the transformed statistic  $\tilde{R}_1^a$ , ( $a = 0.2, 2/3, 1$ ) or  $\tilde{R}_B^0$  is not so different from that of the original power divergence statistic  $R^a$  in the models based on link functions  $g_c(\cdot)$ , ( $c = 0, 1/2, 1$ ) and  $g_P(\cdot)$ .



**Fig. 8.** Simulated average power  $P$  against an alternative model (27) when original test statistic is  $R^{2/3}$  and the model is given by link function  $g_0(\cdot)$ ,  $g_{1/2}(\cdot)$ ,  $g_1(\cdot)$  and  $g_P(\cdot)$  for true parameters (i) and (ii) and sample design (A) with  $n_A = 10$ :  $\circ$ ,  $\diamond$  and  $\triangle$  are the values for  $R^{2/3}$  when  $\varepsilon = 0.01$ , 0.05 and 0.10, respectively, and  $\bullet$ ,  $\blacklozenge$  and  $\blacktriangle$  are the values for  $\tilde{R}_I^{2/3}$  when  $\varepsilon = 0.01$ , 0.05 and 0.10, respectively. The 1st column is for design matrix (I), the 2nd column is for design matrix (II), the 3rd column is for design matrix (III), and the 4th column is for design matrix (IV).

## 5. Real data application

In this section, by applying the transformed statistics to real data, we discuss the importance of the proposed transformed statistics. We use data partly extracted from a case-control study on oesophageal cancer reported by [20] which is referred to in Table 1 of [6]. Data were collected from 200 male cases of oesophageal cancer diagnosed in one of the hospitals in the French department Ile-et-Vilaine (Brittany). The stratum variable age is controlled by defining 10-year age strata. Table 1 shows the number  $n_\alpha$  of persons in each age stratum  $\alpha = 1, \dots, 6$  and the number  $y_\alpha$  of heavy consumers of tobacco ( $\geq 10$  g/day) in each age stratum  $\alpha = 1, \dots, 6$ . As values of a covariate variable, we use 30, 40, 50, 60, 70, and 80 for each age stratum.

The generalized linear model that we use is given by (1) with  $N = 6$ ,  $p = 2$ , and  $x_{\alpha 1} = 1$  and  $x_{\alpha 2} = x_\alpha$ , ( $\alpha = 1, \dots, N$ ), that is,

$$\pi_\alpha^g = g^{-1}(\beta_1 + \beta_2 x_\alpha), \quad (\alpha = 1, \dots, 6).$$

We consider the model when  $g = g_1$ , that is, the logistic regression model and the model when  $g = g_P$ , that is, a probit model. We consider testing the null hypothesis  $H_0^g$  given by (6) using power divergence statistics  $R^0$  (log likelihood ratio statistic),  $R^{0.2}$ ,  $R^{2/3}$ , and  $R^1$  (Pearson's  $X^2$  statistic) at the significance level of 0.05. For each statistic, observed values of the original statistics and the transformed statistics of data in Table 1 for the logistic regression model and for the probit model are shown in Tables 2 and 3, respectively. The nominal critical value of a significance level of 0.05 by using a chi-square distribution is  $\chi_4^2(0.05) = 9.488$ . Therefore, in the logistic regression model and the probit model, if we consider testing  $H_0^g$  using the original power divergence test statistics  $R^0$ ,  $R^{0.2}$ ,  $R^{2/3}$ , and  $R^1$ ,  $H_0^g$  is rejected at the significance level of 0.05 in all tests. On the other hand, if we consider testing  $H_0^g$  using the transformed statistics  $\tilde{R}_B^g$ ,  $\tilde{R}_I^{0.2}$ ,  $\tilde{R}_I^{2/3}$ , and  $\tilde{R}_I^1$ ,  $H_0^g$  is accepted at the significance level of 0.05 in all tests.

We consider the distribution of power divergence statistics  $R^a$ , ( $a = 0, 0.2, 2/3, 1$ ), where statistics are constructed by random variable  $Y_\alpha$ , ( $\alpha = 1, \dots, N$ ), provided that  $Y_\alpha$ , ( $\alpha = 1, \dots, N$ ) is independently distributed according to the binomial distribution  $B(n_\alpha, \hat{\pi}_\alpha^g)$ , ( $\alpha = 1, \dots, N$ ). Let  $R^a(0.05)$ , ( $a = 0, 0.2, 2/3, 1$ ) be the upper 0.05 point of the distribution of  $R^a$ . Then, by using  $R^a(0.05)$ , we can execute an exact test at a significance level of 0.05. Therefore, as an accurate approximation of  $R^a(0.05)$ , we consider a simulated approximation of  $R^a(0.05)$  as follows.

For each  $\alpha$ , by generating  $n_\alpha$ , ( $\alpha = 1, \dots, 6$ ) binomial random numbers that are distributed according to  $B(1, \hat{\pi}_\alpha^g)$ , ( $\alpha = 1, \dots, 6$ ), we obtain  $y_\alpha^{g*}$ , ( $\alpha = 1, \dots, 6$ ), which are observed values of  $Y_\alpha$ , ( $\alpha = 1, \dots, 6$ ). From  $y_\alpha^{g*}$ , ( $\alpha = 1, \dots, 6$ ), we calculate the maximum likelihood estimate of  $\beta$  and observed value of  $R^a$ , ( $a = 0, 0.2, 2/3, 1$ ). By repeating this process

**Table 1**

Age distribution and number of heavy consumers of tobacco, Ille-et-Vilaine study of oesophageal cancer.

Source: Data from [20].

$\alpha$	Age stratum	Covariate ( $x_\alpha$ )	Number of persons ( $n_\alpha$ )	Number of heavy consumers of tobacco ( $y_\alpha$ )
1	25–34	30	1	1
2	35–44	40	9	7
3	45–54	50	46	32
4	55–64	60	76	51
5	65–74	70	55	24
6	75+	80	13	7

$J = 10^5$  times, we obtain  $J$  observed values  $R_{(j)}^a$ , ( $j = 1, \dots, J$ ). By sorting  $R_{(j)}^a$ , ( $j = 1, \dots, J$ ) in large order, we adopt the  $0.05 \times J = 0.5 \times 10^4$  th value as an approximation of  $R^a(0.05)$  and put them  $\tilde{R}_S^a(0.05)$ , ( $a = 0, 0.2, 2/3, 1$ ).

Simulated critical values of significance level of 0.05 for  $R^a$  ( $a = 0, 0.2, 2/3, 1$ ) in the logistic regression model and the probit model are shown in Tables 2 and 3, respectively. From Tables 2 and 3, we find that

$$R_S^a(0.05) > R^a, \quad (a = 0, 0.2, 2/3, 1).$$

Therefore, the results of all of the tests by using the simulated critical value are accepted at the significance level of 0.05. That is, the tests using the nominal critical value lead to a conclusion opposite to that obtained by the tests using the simulated critical value. These results occur on account of pooriness of  $\chi^2$  approximation for the upper probability of the power divergence statistics  $R^a$ , ( $a = 0, 0.2, 2/3, 1$ ).

On the other hand, we calculate the simulated approximation of critical values of significance level of 0.05 for  $\tilde{R}_B^0$  and  $\tilde{R}_I^a$ , ( $a = 0.2, 2/3, 1$ ) of data in Table 1 in the same way as that for  $R_S^a(0.05)$  and put them  $\tilde{R}_S^a(0.05)$ , ( $a = 0, 0.2, 2/3, 1$ ). The values of  $\tilde{R}_S^a(0.05)$  in the logistic regression model and the probit model are also shown in Tables 2 and 3, respectively. From Tables 2 and 3, we find that

$$\tilde{R}_S^0(0.05) > \tilde{R}_B^0,$$

and

$$\tilde{R}_S^a(0.05) > \tilde{R}_I^a, \quad (a = 0.2, 2/3, 1).$$

Therefore, the results of all of the tests by using the simulated critical value are also accepted at the significance level of 0.05. That is, the results of tests using the nominal critical value coincide with the results of tests using the simulated critical value. This is an example of results of a test using the nominal critical value of the proposed transformed statistics  $\tilde{R}_B^0$  and  $\tilde{R}_I^a$ , ( $a = 0.2, 2/3, 1$ ) being more reliable than results of a test using the original power divergence statistics  $R^a$ , ( $a = 0, 0.2, 2/3, 1$ ).

## 6. Concluding remarks

Using the continuous term of the expression of asymptotic expansion for the distribution of statistic  $C_\phi$  based on  $\phi$ -divergence under a null hypothesis, we propose a transformation that improves the speed of convergence to the chi-square limiting distribution of  $C_\phi$ . In the case of power divergence statistic  $R^a$ , numerical comparison shows that the transformed power divergence statistic is effective for improving the speed of convergence to the chi-square limiting distribution without decreasing power in many models including the logistic regression model and probit model. This improvement increases the reliability of the results of an asymptotic test.

Pardo and Pardo [13] considered the problem of goodness of fit to generalized linear models with binary data using the minimum  $\phi$ -divergence estimator instead of the maximum likelihood estimator. Test statistic  $C_\phi$  in this paper is the statistic using MLE of  $\beta$ . Let the statistic defined by the minimum  $\phi^*$ -divergence estimator of  $\beta$  instead of MLE of  $\beta$  be  $C_{\phi, \phi^*}$ , then  $C_\phi = C_{\phi, \phi_0}$ . So, the family of the test statistics  $C_{\phi, \phi^*}$  includes the family of test statistics  $C_\phi$  as a special case. Therefore, it is not possible to apply results presented in this paper immediately to the statistics  $C_{\phi, \phi^*}$ . However, by using an expansion of the minimum  $\phi^*$ -divergence estimator instead of expansion of MLE, we can derive an expansion of test statistics  $C_{\phi, \phi^*}$ . Then, by using the expansion of  $C_{\phi, \phi^*}$ , it is possible to construct new transformed statistics based on  $C_{\phi, \phi^*}$ .

## Acknowledgments

The authors are very grateful to the reviewers for their valuable comments and suggestions. This research is partially supported by the Grants-in-aid for Scientific Research of Japan Society for the Promotion of Science (C) 24540133.

**Table 2**

Observed values and simulated critical values for significance level of 0.05 of original power divergence statistics and transformed statistics in the logistic regression model.

$a$	$R^a$	$\tilde{R}_l^a (a \neq 0), \tilde{R}_g^0$		
	Observed value of test statistic	Simulated critical value $\tilde{R}_S^a(0.05)$	Observed value of test statistic	Simulated critical value $\tilde{R}_S^a(0.05)$
0.0	12.175	15.148	2.747	10.074
0.2	11.819	14.943	7.080	12.883
2/3	11.355	14.768	5.891	12.903
1.0	11.191	14.790	3.661	11.296

**Table 3**

Observed values and simulated critical values for significance level of 0.05 of original power divergence statistics and transformed statistics in the probit model.

$a$	$R^a$	$\tilde{R}_l^a (a \neq 0), \tilde{R}_g^0$		
	Observed value of test statistic	Simulated critical value $\tilde{R}_S^a(0.05)$	Observed value of test statistic	Simulated critical value $\tilde{R}_S^a(0.05)$
0.0	12.207	15.174	0.279	9.900
0.2	11.871	14.986	5.447	12.743
2/3	11.430	14.803	3.767	12.752
1.0	11.272	14.827	2.163	11.177

## Appendix A. Proof of Theorem 1

By transformation (10), statistic  $C_\phi$  can be rewritten as

$$C_\phi(\mathbf{W}) = 2 \sum_{\alpha=1}^N n_\alpha \left\{ \hat{\pi}_\alpha^g(\mathbf{W}) \phi \left( \frac{\pi_\alpha^g + W_\alpha(\sqrt{n_\alpha})^{-1}}{\hat{\pi}_\alpha^g(\mathbf{W})} \right) + (1 - \hat{\pi}_\alpha^g(\mathbf{W})) \phi \left( \frac{1 - \pi_\alpha^g - W_\alpha(\sqrt{n_\alpha})^{-1}}{1 - \hat{\pi}_\alpha^g(\mathbf{W})} \right) \right\}.$$

If we regard

$$h^g(\mathbf{w}) \left\{ 1 + \frac{1}{\sqrt{n}} h_1^g(\mathbf{w}) + \frac{1}{n} h_2^g(\mathbf{w}) + \frac{1}{n\sqrt{n}} h_3^g(\mathbf{w}) \right\}$$

as the continuous density function of  $\mathbf{W}$ , then we can regard

$$J_1^{g,\phi}(x) = \int \cdots \int_{U_\phi^g(x)} h^g(\mathbf{w}) \left\{ 1 + \frac{1}{\sqrt{n}} h_1^g(\mathbf{w}) + \frac{1}{n} h_2^g(\mathbf{w}) + \frac{1}{n\sqrt{n}} h_3^g(\mathbf{w}) \right\} d\mathbf{w}$$

as the distribution function of  $C_\phi(\mathbf{W})$ , where  $U_\phi^g(x)$  is defined by (15). So, the characteristic function of  $C_\phi(\mathbf{W})$  is calculated as

$$\psi_\phi^g(u) = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} [\exp\{iuC_\phi(\mathbf{w})\}] h^g(\mathbf{w}) \left\{ 1 + \frac{1}{\sqrt{n}} h_1^g(\mathbf{w}) + \frac{1}{n} h_2^g(\mathbf{w}) + \frac{1}{n\sqrt{n}} h_3^g(\mathbf{w}) \right\} d\mathbf{w}. \quad (\text{A.1})$$

We can expand  $C_\phi(\mathbf{w})$  as

$$C_\phi(\mathbf{w}) = \tau_0^g(\mathbf{w}) + \frac{1}{\sqrt{n}} \tau_1^{g,\phi}(\mathbf{w}) + \frac{1}{n} \tau_2^{g,\phi}(\mathbf{w}) + \frac{1}{n\sqrt{n}} \tau_3^{g,\phi}(\mathbf{w}) + O(n^{-2}), \quad (\text{A.2})$$

where

$$\tau_0^g(\mathbf{w}) = \mathbf{w}'(\Omega^{-1} - \Xi)\mathbf{w},$$

$\Xi = (\xi_{\alpha\beta})$  is a  $N \times N$  matrix,

$$\xi_{\alpha\beta} = \frac{\sqrt{\mu_\alpha} G_1(\alpha)}{\pi_\alpha^g(1 - \pi_\alpha^g)} \frac{\sqrt{\mu_\beta} G_1(\beta)}{\pi_\beta^g(1 - \pi_\beta^g)} \sigma_{\alpha\beta}, \quad (\alpha, \beta = 1, \dots, N),$$

$$\tau_1^{g,\phi}(\mathbf{w}) = \sum_{a=0}^3 \left( \sum_{\alpha=1}^N B_{a+1}^1(\alpha) C_{1(\alpha)}(\mathbf{w})^{3-a} w_\alpha^a \right),$$

$$\begin{aligned} \tau_2^{g,\phi}(\mathbf{w}) &= \sum_{a=0}^2 \left( \sum_{\alpha=1}^N B_{a+1}^2(\alpha) C_{2(\alpha)}(\mathbf{w})^{2-a} C_{1(\alpha)}(\mathbf{w})^{2a} \right) + \sum_{a=0}^1 \left( \sum_{\alpha=1}^N B_{a+4}^2(\alpha) C_{2(\alpha)}(\mathbf{w})^{1-a} C_{1(\alpha)}(\mathbf{w})^{1+2a} w_\alpha \right) \\ &\quad + \sum_{a=0}^1 \left( \sum_{\alpha=1}^N B_{a+6}^2(\alpha) C_{2(\alpha)}(\mathbf{w})^{1-a} C_{1(\alpha)}(\mathbf{w})^{2a} w_\alpha^2 \right) + \sum_{\alpha=1}^N B_8^2(\alpha) C_{1(\alpha)}(\mathbf{w}) w_\alpha^3 + \sum_{\alpha=1}^N B_9^2(\alpha) w_\alpha^4, \end{aligned}$$



$$\begin{aligned}
B_1^1(\alpha) &= \frac{\mu_\alpha}{3(\pi_\alpha^g)^2(1-\pi_\alpha^g)^2} \{3\pi_\alpha^g(1-\pi_\alpha^g)G_1(\alpha)G_2(\alpha) - (3+\phi'''(1))(1-2\pi_\alpha^g)G_1(\alpha)^3\}, \\
B_2^1(\alpha) &= \frac{\sqrt{\mu_\alpha}}{(\pi_\alpha^g)^2(1-\pi_\alpha^g)^2} \{-\pi_\alpha^g(1-\pi_\alpha^g)G_2(\alpha) + (2+\phi'''(1))(1-2\pi_\alpha^g)G_1(\alpha)^2\}, \\
B_3^1(\alpha) &= -\frac{(1+\phi'''(1))(1-2\pi_\alpha^g)G_1(\alpha)}{(\pi_\alpha^g)^2(1-\pi_\alpha^g)^2}, \quad B_4^1(\alpha) = \frac{\phi'''(1)(1-2\pi_\alpha^g)}{3\sqrt{\mu_\alpha}(\pi_\alpha^g)^2(1-\pi_\alpha^g)^2}, \\
B_1^2(\alpha) &= \frac{\mu_\alpha G_1(\alpha)^2}{\pi_\alpha^g(1-\pi_\alpha^g)}, \quad B_2^2(\alpha) = 3B_1^1(\alpha), \\
B_3^2(\alpha) &= \frac{\mu_\alpha}{12(\pi_\alpha^g)^3(1-\pi_\alpha^g)^3} \{(\pi_\alpha^g)^2(1-\pi_\alpha^g)^2(3G_2(\alpha)^2 + 4G_1(\alpha)G_3(\alpha)) - 6(3+\phi'''(1))\pi_\alpha^g \\
&\quad \times (1-\pi_\alpha^g)(1-2\pi_\alpha^g)G_1(\alpha)^2G_2(\alpha) + (12+8\phi'''(1)+\phi^{(4)}(1))(1-3\pi_\alpha^g+3(\pi_\alpha^g)^2)G_1(\alpha)^4\}, \\
B_4^2(\alpha) &= 2B_2^1(\alpha), \\
B_5^2(\alpha) &= \frac{\sqrt{\mu_\alpha}}{3(\pi_\alpha^g)^3(1-\pi_\alpha^g)^3} \{-(\pi_\alpha^g)^2(1-\pi_\alpha^g)^2G_3(\alpha) + 3(2+\phi'''(1))\pi_\alpha^g(1-\pi_\alpha^g) \\
&\quad \times (1-2\pi_\alpha^g)G_1(\alpha)G_2(\alpha) - (6+6\phi'''(1)+\phi^{(4)}(1))(1-3\pi_\alpha^g+3(\pi_\alpha^g)^2)G_1(\alpha)^3\}, \\
B_6^2(\alpha) &= B_3^1(\alpha), \\
B_7^2(\alpha) &= \frac{1}{2(\pi_\alpha^g)^3(1-\pi_\alpha^g)^3} \{-(1+\phi'''(1))\pi_\alpha^g(1-\pi_\alpha^g)(1-2\pi_\alpha^g)G_2(\alpha) \\
&\quad + (2+4\phi'''(1)+\phi^{(4)}(1))(1-3\pi_\alpha^g+3(\pi_\alpha^g)^2)G_1(\alpha)^2\}, \\
B_8^2(\alpha) &= -\frac{(2\phi'''(1)+\phi^{(4)}(1))(1-3\pi_\alpha^g+3(\pi_\alpha^g)^2)G_1(\alpha)}{3\sqrt{\mu_\alpha}(\pi_\alpha^g)^3(1-\pi_\alpha^g)^3}, \\
B_9^2(\alpha) &= \frac{\phi^{(4)}(1)(1-3\pi_\alpha^g+3(\pi_\alpha^g)^2)}{12\mu_\alpha(\pi_\alpha^g)^3(1-\pi_\alpha^g)^3}, \\
C_{1(\alpha)}(\mathbf{w}) &= \sum_{m=1}^p x_{\alpha m} \left( \sum_{k=1}^p \kappa^{m,k} M_k(\mathbf{w}) \right), \quad (\alpha = 1, \dots, N), \\
C_{2(\alpha)}(\mathbf{w}) &= \sum_{m=1}^p x_{\alpha m} \left\{ \sum_{k=1}^p M^{m,k}(\mathbf{w}) M_k(\mathbf{w}) \right. \\
&\quad \left. + \frac{1}{2} \sum_{k_1=1}^p \dots \sum_{k_5=1}^p \kappa^{m,k_3} \kappa^{k_1,k_4} \kappa^{k_2,k_5} \kappa_{k_3,k_4,k_5} M_{k_1}(\mathbf{w}) M_{k_2}(\mathbf{w}) \right\}, \quad (\alpha = 1, \dots, N), \\
M_k(\mathbf{w}) &= \sum_{\lambda=1}^N \sqrt{\mu_\lambda} x_{\lambda k} G_1(\lambda) \{\pi_\lambda^g(1-\pi_\lambda^g)\}^{-1} w_\lambda, \quad (k = 1, \dots, p), \\
J_{i,j}(\mathbf{w}) &= \sum_{\lambda=1}^N \sqrt{\mu_\lambda} x_{\lambda i} x_{\lambda j} \left\{ -\frac{(1-2\pi_\lambda^g)G_1(\lambda)^2}{(\pi_\lambda^g)^2(1-\pi_\lambda^g)^2} + \frac{G_2(\lambda)}{\pi_\lambda^g(1-\pi_\lambda^g)} \right\} w_\lambda, \quad (i, j = 1, \dots, p), \\
\kappa_{i,j,k} &= \sum_{\lambda=1}^N \mu_\lambda x_{\lambda i} x_{\lambda j} x_{\lambda k} \left\{ 2\frac{(1-2\pi_\lambda^g)G_1(\lambda)^3}{(\pi_\lambda^g)^2(1-\pi_\lambda^g)^2} - 3\frac{G_1(\lambda)G_2(\lambda)}{\pi_\lambda^g(1-\pi_\lambda^g)} \right\}, \quad (i, j, k = 1, \dots, p), \\
G_i(\alpha) &= u^{(i)}(\mathbf{x}'_\alpha \boldsymbol{\beta}), \quad (\alpha = 1, \dots, N, i = 1, 2, 3),
\end{aligned}$$

$J(\mathbf{w}) = (J_{i,j}(\mathbf{w}))$  is a  $p \times p$  matrix,  $M^{i,j}(\mathbf{w})$  is the  $(i, j)$ -element of matrix  $K^{-1}J(\mathbf{w})K^{-1}$ ,  $\Omega$  is defined by (12),  $\sigma_{\alpha\beta}$  and  $K^{-1} = (\kappa^{i,j})$  are defined in Theorem 1, and  $\tau_3^{g,\phi}(\mathbf{w})$  is a homogeneous polynomial of degree 5 with respect to variables  $w_1, \dots, w_N$ . Then, from (11), (A.1) and (A.2), we obtain

$$\begin{aligned}
\psi_\phi^g(u) &= (1-2iu)^{-(N-p)/2} \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} (2\pi)^{-N/2} |\Lambda|^{-1/2} \left\{ \exp\left(-\frac{1}{2}\mathbf{w}'\Lambda^{-1}\mathbf{w}\right) \right\} \\
&\quad \times \{1 + n^{-1/2}D_1(\mathbf{w}) + n^{-1}D_2(\mathbf{w}) + n^{-3/2}D_3(\mathbf{w})\} d\mathbf{w} + O(n^{-2}),
\end{aligned} \tag{A.3}$$

where  $\Lambda = (1-2iu)^{-1}(\Omega - 2iu\Omega\Xi\Omega)$ ,

$$D_1(\mathbf{w}) = h_1^g(\mathbf{w}) + (iu)\tau_1^{g,\phi}(\mathbf{w}),$$

$$D_2(\mathbf{w}) = h_2^g(\mathbf{w}) + (iu)\tau_1^{g,\phi}(\mathbf{w})h_1^g(\mathbf{w}) + (iu)\tau_2^{g,\phi}(\mathbf{w}) + \frac{1}{2}(iu)^2 \left\{ \tau_1^{g,\phi}(\mathbf{w}) \right\}^2,$$

and degrees of all terms of polynomial  $D_3(\mathbf{w})$  are odd. Therefore, by carrying out the integration of (A.3), the characteristic function  $\psi_\phi^g(u)$  is expanded as

$$\psi_\phi^g(u) = (1 - 2iu)^{-(N-p)/2} \left[ 1 + \frac{1}{n} \sum_{j=0}^3 (1 - 2iu)^{-j} v_j^{g,\phi} + O(n^{-2}) \right]. \quad (\text{A.4})$$

By inverting (A.4), we obtain (13). We have completed the proof of Theorem 1.

## Appendix B. Proof of Theorem 2

The function  $S_1(\sqrt{n}w_\gamma + n\pi_\gamma^g)$  defined by (19) is differentiable except when  $w_\gamma \in L_\gamma$  defined by (18), and  $h^g(\mathbf{w})$  is a differentiable function on  $R_N$ . Therefore, by (11), (12) and (19), we obtain

$$\begin{aligned} [S_1(\sqrt{n}w_\gamma + n\pi_\gamma^g) h^g(\mathbf{w})]_{\eta_\gamma^{g,\phi}(\tilde{\mathbf{w}}_\gamma)}^{\theta_\gamma^{g,\phi}(\tilde{\mathbf{w}}_\gamma)} &= \left\{ (2\pi)^N \prod_{\alpha=1}^N \pi_\alpha^g (1 - \pi_\alpha^g) \right\}^{-1/2} \\ &\times \left\{ \sqrt{n} \int_{\eta_\gamma^{g,\phi}(\tilde{\mathbf{w}}_\gamma)}^{\theta_\gamma^{g,\phi}(\tilde{\mathbf{w}}_\gamma)} \exp\left(-\frac{1}{2} \mathbf{w}' \Omega^{-1} \mathbf{w}\right) dw_\gamma - \sum_{\substack{w_\gamma = \eta_\gamma^{g,\phi}(\tilde{\mathbf{w}}_\gamma) \\ w_\gamma \in L_\gamma}}^{\theta_\gamma^{g,\phi}(\tilde{\mathbf{w}}_\gamma)} \exp\left(-\frac{1}{2} \mathbf{w}' \Omega^{-1} \mathbf{w}\right) \right. \\ &\left. - \frac{1}{\pi_\gamma^g (1 - \pi_\gamma^g)} \int_{\eta_\gamma^{g,\phi}(\tilde{\mathbf{w}}_\gamma)}^{\theta_\gamma^{g,\phi}(\tilde{\mathbf{w}}_\gamma)} w_\gamma S_1(\sqrt{n}w_\gamma + n\pi_\gamma^g) \exp\left(-\frac{1}{2} \mathbf{w}' \Omega^{-1} \mathbf{w}\right) dw_\gamma \right\}. \end{aligned} \quad (\text{B.1})$$

Then from (17) and (B.1), we obtain the following:

$$J_2^{g,\phi}(x) = \left\{ (2\pi)^N \prod_{\alpha=1}^N \pi_\alpha^g (1 - \pi_\alpha^g) \right\}^{-1/2} (\Theta_1^{g,\phi} + \Theta_2^{g,\phi*} - \Theta_3^{g,\phi*},$$

where

$$\Theta_3^{g,\phi*} = \int \dots \int_{U_\phi^g(x)} h^g(\mathbf{w}) d\mathbf{w}$$

and

$$\begin{aligned} \Theta_2^{g,\phi*} &= n^{-N/2} \sum_{\gamma=1}^N n^{(\gamma-1)/2} \frac{1}{\pi_\gamma^g (1 - \pi_\gamma^g)} \sum_{w_{\gamma+1} \in L_{\gamma+1}} \dots \sum_{w_N \in L_N} \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} \chi_{U_\gamma^{g,\phi}(\tilde{\mathbf{w}}_\gamma)} \\ &\times \int_{\eta_\gamma^{g,\phi}(\tilde{\mathbf{w}}_\gamma)}^{\theta_\gamma^{g,\phi}(\tilde{\mathbf{w}}_\gamma)} w_\gamma S_1(\sqrt{n}w_\gamma + n\pi_\gamma^g) \exp\left(-\frac{1}{2} \mathbf{w}' \Omega^{-1} \mathbf{w}\right) dw_1 \dots dw_\gamma. \end{aligned}$$

If we regard  $h^g(\mathbf{w})$  as the density function of  $\mathbf{W}$ , then we can regard  $\Theta_3^{g,\phi*}$  as the distribution function of  $C_\phi(\mathbf{W})$ . Then, by expanding the characteristic function of  $C_\phi(\mathbf{W})$  and inverting it, we can approximate  $\Theta_3^{g,\phi*} = \Theta_3^{g,\phi} + O(n^{-2})$ . Furthermore, we can approximate  $\Theta_2^{g,\phi*} = \Theta_2^{g,\phi} + O(n^{-2})$ . Therefore, we obtain (20). We have completed the proof of Theorem 2.

## References

- [1] F.J. Aranda-Ordaz, On two families of transformations to additivity for binary response data, *Biometrika* 68 (1981) 357–363.
- [2] O.E. Barndorff-Nielsen, D.R. Cox, Bartlett adjustments to the likelihood ratio statistic and the distribution of maximum likelihood estimator, *J. R. Stat. Soc. Ser. B* 46 (1984) 483–495.
- [3] O.E. Barndorff-Nielsen, P. Hall, On the level-error after Bartlett adjustment of the likelihood ratio statistic, *Biometrika* 75 (1988) 374–378.
- [4] R.N. Bhattacharya, R. Ranga Rao, Normal Approximation and Asymptotic Expansion, Wiley, New York, 1976.
- [5] N. Cressie, T.R.C. Read, Multinomial goodness-of-fit tests, *J. R. Stat. Soc. Ser. B* 46 (1984) 440–464.
- [6] K. Drescher, W. Schill, Attributable risk estimation from case-control data via logistic regression, *Biometrics* 47 (1991) 1247–1256.
- [7] Y. Fujikoshi, Transformations with improved chi-squared approximations, *J. Multivariate Anal.* 72 (2000) 249–263.
- [8] Y. Kakizawa, Higher order monotone Bartlett-type adjustment for some multivariate test statistics, *Biometrika* 83 (1996) 923–927.
- [9] D.N. Lawley, A general method for approximating to the distribution of the likelihood ratio criteria, *Biometrika* 43 (1956) 295–303.
- [10] M.L. Menéndez, J.A. Pardo, L. Pardo, M.C. Pardo, Asymptotic approximations for the distributions of the  $(h, \phi)$ -divergence goodness-of-fit statistics: application to Renyi's statistic, *Kybernetes* 26 (4) (1997) 442–452.
- [11] J.A. Nelder, R.W.M. Wedderburn, Generalized linear models, *J. R. Stat. Soc. Ser. A* 135 (1972) 370–384.

- [12] Pan Ei Htwe, N. Taneichi, Y. Sekiya, Improved approximations for the distributions of multinomial goodness-of-fit statistics based on  $\phi$ -divergence under nonlocal alternatives, *J. Japan Statist. Soc.* 41 (2) (2011) 121–142.
- [13] J.A. Pardo, M.C. Pardo, Minimum  $\phi$ -divergence estimator and  $\phi$ -divergence statistics in generalized linear models with binary data, *Methodol. Comput. Appl. Probab.* 10 (2008) 357–379.
- [14] T.R.C. Read, Closer asymptotic approximations for the distributions of the power divergence goodness-of-fit statistics, *Ann. Inst. Statist. Math.* 36 (1984) 59–69.
- [15] Y. Sekiya, N. Taneichi, Improvement of approximations for the distributions of multinomial goodness-of-fit statistics under nonlocal alternatives, *J. Multivariate Anal.* 91 (2004) 199–223.
- [16] M. Siotani, Y. Fujikoshi, Asymptotic approximations for the distributions of multinomial goodness-of-fit statistics, *Hiroshima Math. J.* 14 (1984) 115–124.
- [17] N. Taneichi, Y. Sekiya, A. Suzukawa, 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.
- [18] N. Taneichi, Y. Sekiya, A. Suzukawa, Asymptotic approximations for the distributions of the multinomial goodness-of-fit statistics under local alternatives, *J. Multivariate Anal.* 81 (2002) 335–359.
- [19] N. Taneichi, Y. Sekiya, J. Toyama, Improved transformed deviance statistic for testing a logistic regression model, *J. Multivariate Anal.* 102 (2011) 1263–1279.
- [20] A.J. Tuyns, G. Péquignot, O.M. Jensen, Le cancer de l'oesophage en Ile-et-Vilaine en fonction des niveaux de consommation d'alcool et tabac, *Bull. Cancer (Paris)* 64 (1977) 45–60.
- [21] H. Yanagihara, Transformations for improving normal and chi-squared approximations, Master's Thesis, Hiroshima University, 1998 (in Japanese).
- [22] J.K. Yarnold, Asymptotic approximations for the probability that a sum of lattice random vectors lies in a convex set, *Ann. Math. Statist.* 43 (1972) 1566–1580.