SUT Journal of Mathematics
Vol. 52, No. 2 (2016), 193–214

# Improved transformation of $\phi$-divergence goodness-of-fit test statistics based on minimum $\phi^*$-divergence estimator for GLIM of binary data

**Nobuhiro Taneichi, Yuri Sekiya and Jun Toyama**

(Received September 16, 2016; Revised November 14, 2016)

**Abstract.** Generalized linear models of binary data including a logistic regression model and a probit model are considered. For testing the null hypothesis that the considered model is correct, the $\phi$-divergence family of goodness-of-fit test statistics $C_{\phi\phi^*}$ that is based on a minimum $\phi^*$-divergence estimator is considered. The family of statistics $C_{\phi\phi^*}$ includes a power divergence family of statistics $R^{a,b}$ that is based on a minimum power divergence estimator. The derivation of an expression of a continuous term of asymptotic expansion for the distribution of $C_{\phi\phi^*}$ under the null hypothesis is shown. Using the expression, a transformed $C_{\phi\phi^*}$ statistic that improves the speed of convergence to the chi-square limiting distribution of $C_{\phi\phi^*}$ is obtained. In the case of $R^{a,b}$, it is numerically shown that the transformed statistics usually perform better than the original statistics with respect to speed of convergence to the chi-square limiting distribution and it is also numerically shown that the power of the transformed statistics is almost the same as that of the original statistics.

*AMS* 2010 *Mathematics Subject Classification.* 62E20, 62H10.

*Key words and phrases.* Asymptotic expansion, binary data, $\phi$-divergence statistics, generalized linear model, improved transformation, minimum $\phi^*$-divergence estimator.

## §1. Introduction

We discuss generalized linear models (Nelder and Wedderburn [10]) 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 binomial distributions $B(n_\alpha, \pi_\alpha)$, $\alpha = 1, \ldots, N$. If we use a monotone and differentiable function $g$ as a link function, we obtain a generalized linear model for binary data as follows.

$$g(\pi_\alpha) = \boldsymbol{x}'_\alpha \boldsymbol{\beta} \ (\alpha = 1, \ldots, N), \tag{1.1}$$

where $\boldsymbol{x}_\alpha = (x_{\alpha 1}, \ldots, x_{\alpha p})'(\alpha = 1, \ldots, N)$ are covariate vectors and $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_p)'$ is an unknown parameter vector and $p < N$. We consider a minimum $\phi^*$-divergence estimator of model (1.1) and also consider a $\phi$-divergence goodness-of-fit test statistic based on the estimator. Let $y_\alpha$ $(\alpha = 1, \ldots, N)$ be an observed value of $Y_\alpha$ $(\alpha = 1, \ldots, N)$, then the minimum $\phi^*$-divergence estimator of model (1.1) is given by

$$\hat{\boldsymbol{\beta}}^{g\phi^*} = \arg\min_{\beta \in \Theta} D_{\phi^*},$$

where

$$D_{\phi^*} = \frac{1}{N} \sum_{\alpha=1}^{N} n_\alpha \left\{ \pi_\alpha(\boldsymbol{\beta}) \phi^* \left( \frac{\dfrac{y_\alpha}{n_\alpha}}{\pi_\alpha(\boldsymbol{\beta})} \right) + (1 - \pi_\alpha(\boldsymbol{\beta})) \phi^* \left( \frac{1 - \dfrac{y_\alpha}{n_\alpha}}{1 - \pi_\alpha(\boldsymbol{\beta})} \right) \right\},$$

where $\phi^*$ is a real convex function in $(0, \infty)$ satisfying $\phi^*(1) = {\phi^*}'(1) = 0$, ${\phi^*}''(1) = 1$, $0\phi^*(0/0) = 0$, $0\phi^*(x/0) = \lim_{u \to \infty} \phi^*(u)/u$, and $\Theta$ is an open subset of $R^p$ (Pardo [11]). When we choose a convex function

$$\phi_a(t) = \begin{cases} \{a(a+1)\}^{-1}\{t^{a+1} - t + a(1-t)\} & (a \neq 0, -1) \\ t \log t + 1 - t & (a = 0) \\ -\log t - 1 + t & (a = -1), \end{cases} \tag{1.2}$$

as $\phi^*(t)$, $D_{\phi_0}$ becomes a Kullback divergence measure (Kullback [7]). Then, in this case, estimator $\hat{\boldsymbol{\beta}}^{g\phi_0}$ becomes the maximum likelihood estimator. Therefore, the maximum likelihood estimator is a special case of the minimum $\phi^*$-divergence estimator.

In order to test the null hypothesis

$$H_0^g : \pi_\alpha = \pi_\alpha(\boldsymbol{\beta}) = g^{-1}(\boldsymbol{x}'_\alpha \boldsymbol{\beta}) \ (\alpha = 1, \ldots, N), \tag{1.3}$$

we consider the family of $\phi$-divergence statistics based on the minimum $\phi^*$-divergence estimator

$$C_{\phi\phi^*} = 2 \sum_{\alpha=1}^{N} n_\alpha \left\{ \hat{\pi}_\alpha^{g\phi^*} \phi \left( \frac{\dfrac{Y_\alpha}{n_\alpha}}{\hat{\pi}_\alpha^{g\phi^*}} \right) + (1 - \hat{\pi}_\alpha^{g\phi^*}) \phi \left( \frac{1 - \dfrac{Y_\alpha}{n_\alpha}}{1 - \hat{\pi}_\alpha^{g\phi^*}} \right) \right\}, \tag{1.4}$$

where $\hat{\pi}_\alpha^{g\phi^*} = \pi_\alpha(\hat{\boldsymbol{\beta}}^{g\phi^*})$ $(\alpha = 1, \ldots, N)$, $\hat{\boldsymbol{\beta}}^{g\phi^*} = (\hat{\beta}_1^{g\phi^*}, \ldots, \hat{\beta}_p^{g\phi^*})'$ is the minimum $\phi^*$-divergence estimator of $\boldsymbol{\beta}$ under $H_0^g$ given by (1.3) and $\phi$ satisfies the same conditions of $\phi^*$ (Pardo [11], Pardo and Pardo [12]). The test statistic $C_\phi$ given by (7) in Taneichi *et al.* [21] is written as $C_\phi \equiv C_{\phi\phi_0}$, and therefore the family of statistics given by (1.4) includes that of $C_\phi$.

When we choose convex functions $\phi_a$ and $\phi_b$ given by (1.2) as $\phi$ and $\phi^*$, respectively, in (1.4), $C_{\phi_a\phi_b}$ becomes a power divergence statistic

$$R^{a,b} = 2\sum_{\alpha=1}^{N} n_\alpha \left\{ I^a\left(\frac{Y_\alpha}{n_\alpha},\ \hat{\pi}_\alpha^{g\phi_b}\right) + I^a\left(1-\frac{Y_\alpha}{n_\alpha},\ 1-\hat{\pi}_\alpha^{g\phi_b}\right)\right\}, \tag{1.5}$$

where

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

which is based on the minimum power divergence estimator (Cressie and Read [4], Read and Cressie [14]). Under $H_0^g$, all members of the class of statistics $C_{\phi\phi^*}$ have a $\chi^2_{N-p}$ limiting distribution, assuming the condition that

$$n_\alpha/n \to \mu_\alpha \ (\alpha = 1,\ldots,N) \text{ as } \quad n\to\infty, \tag{1.6}$$

where $n = \sum_{\alpha=1}^N n_\alpha$, $0 < \mu_\alpha < 1$ $(\alpha = 1,\ldots,N)$ and $\sum_{\alpha=1}^N \mu_\alpha = 1$. Using the results, we can use $C_{\phi\phi^*}$ as a goodness-of-fit test statistic for model (1.1).

With regard to the goodness-of-fit test for a multinomial distribution, Yarnold [23] obtained an approximation based on asymptotic expansion for the null distribution of Pearson's $X^2$ statistic. The expansion consists of a term of multivariate Edgeworth expansion for a continuous distribution and a discontinuous term. In a fashion similar to that for Pearson's $X^2$ statistic, approximations based on asymptotic expansions for null distributions of some kinds of multinomial goodness-of-fit statistics have been investigated by Siotani and Fujikoshi [16], Read [13] and Menéndez *et al.* [9]. Edgeworth approximations of the distributions of some kinds of multinomial goodness-of-fit statistics under alternative hypotheses have also been investigated by Taneichi *et al.* [17, 18], and Sekiya and Taneichi [15]. Taneichi and Sekiya [19] discussed approximations for the distribution of $\phi$-divergence statistics for the test of independence in $r \times s$ contingency tables. By using the above theory of approximation, Taneichi *et al.* [21] considered a family of $\phi$-divergence statistics using the maximum likelihood estimator $C_\phi \equiv C_{\phi\phi_0}$ and investigated asymptotic approximation of the distribution of statistics for testing the null hypothesis $H_0^g$ given by (1.3). They proposed transformed $C_\phi$ statistics that improve the speed of convergence to a chi-square limiting distribution.

In this paper, we generalize the family of statistics $C_\phi \equiv C_{\phi\phi_0}$ based on $\phi$-divergence to $C_{\phi\phi^*}$ and investigate an asymptotic approximation of the distribution of $C_{\phi\phi^*}$ under $H_0^g$. Also, we propose transformed $C_{\phi\phi^*}$ statistics.

In Section 2, we first describe a local Edgeworth approximation for the probability of $Y_\alpha$ ($\alpha = 1, \ldots, N$) under $H_0^g$. Next, we consider an expression of asymptotic expansion for the distribution of $C_{\phi\phi^*}$ under $H_0^g$. Evaluation for the continuous term 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 $C_{\phi\phi^*}$ under $H_0^g$. In Section 4, in the case of $R^{a,b}$, performance of the transformed statistic and that of the original statistic are compared numerically.

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

First, we consider a local Edgeworth approximation for the probability of $Y_\alpha$ ($\alpha = 1, \ldots, N$) under null hypothesis $H_0^g$ given by (1.3). Let $Y_\alpha$, $\alpha = 1, \ldots, N$ be distributed according to a binomial distribution $B(n_\alpha, \pi_\alpha^g)$ $\alpha = 1, \ldots, N$, where each $\pi_\alpha^g$ ($\alpha = 1, \ldots, N$) is represented as $\pi_\alpha^g = g^{-1}(\boldsymbol{x}'_\alpha \boldsymbol{\beta})$ ($\alpha = 1, \ldots, N$) by using covariate vectors $\boldsymbol{x}_\alpha = (x_{\alpha 1}, \ldots, x_{\alpha p})'$ and an unknown parameter vector $\boldsymbol{\beta}$. Let

$$W_\alpha = \frac{Y_\alpha - n_\alpha \pi_\alpha^g}{\sqrt{n_\alpha}} \quad (\alpha = 1, \ldots, N). \tag{2.1}$$

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

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

where

$$M = \Big\{ \boldsymbol{y} = (y_1, \ldots, y_N)' : y_1, \ldots, y_N \text{ are non-negative integers that satisfy } y_\alpha \le n_\alpha \ (\alpha = 1, \ldots, N) \Big\}.$$

If we consider only for a limiting distribution of $C_{\phi\phi^*}$, we can discuss under the assumption given by (1.6). In this section, since we consider asymptotic expansion of the distribution of $C_{\phi\phi^*}$, we need an assumption that states the way of converging $n_\alpha/n$ to $\mu_\alpha$ more strictly than the assumption given by (1.6). Therefore, we consider the following Assumption 2.1 instead of the assumption given by (1.6).

**Assumption 2.1.** *$n_\alpha \to \infty$ $(\alpha = 1, \ldots, N)$, as $n \to \infty$, with $n_\alpha$ depending on $n$ in such a way that $n_\alpha/n = \mu_\alpha$ $(\alpha = 1, \ldots, N)$, where $0 < \mu_\alpha < 1$ $(\alpha = 1, \ldots, N)$ and $\sum_{\alpha=1}^{N} \mu_\alpha = 1$.*

With regard to a local Edgeworth approximation for the probability of $Y_\alpha$ $(\alpha = 1, \ldots, N)$ under $H_0^g$, the following lemma is shown in Taneichi *et al.* [21].

**Lemma 2.1.** *For each $\boldsymbol{y} = (y_1, \ldots, y_N)' \in M$, let $\boldsymbol{w} = (w_1, \ldots, w_N)'$, where $w_\alpha = (y_\alpha - n_\alpha \pi_\alpha^g)/\sqrt{n_\alpha}$ $(\alpha = 1, \ldots, N)$. Then, under Assumption 2.1,*

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

*where*

$$h^g(\boldsymbol{w}) = (2\pi)^{-N/2} |\Omega|^{-1/2} \exp\left( -\frac{1}{2} \boldsymbol{w}' \Omega^{-1} \boldsymbol{w} \right), \tag{2.2}$$

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

$$\begin{aligned} h_2^g(\boldsymbol{w}) &= \frac{1}{2} \{h_1^g(\boldsymbol{w})\}^2 - \frac{1}{12} \sum_{\alpha=1}^{N} \frac{1}{\mu_\alpha} \frac{1 - \pi_\alpha^g + (\pi_\alpha^g)^2}{\pi_\alpha^g (1 - \pi_\alpha^g)} \\ &\quad + \frac{1}{4} \sum_{\alpha=1}^{N} \frac{1}{\mu_\alpha} \frac{1 - 2\pi_\alpha^g + 2(\pi_\alpha^g)^2}{(\pi_\alpha^g)^2 (1 - \pi_\alpha^g)^2} w_\alpha^2 \\ &\quad - \frac{1}{12} \sum_{\alpha=1}^{N} \frac{1}{\mu_\alpha} \frac{1 - 3\pi_\alpha^g + 3(\pi_\alpha^g)^2}{(\pi_\alpha^g)^3 (1 - \pi_\alpha^g)^3} w_\alpha^4, \end{aligned}$$

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

*and*

$$\Omega = diag(\pi_1^g(1-\pi_1^g), \ldots, \pi_N^g(1-\pi_N^g)). \tag{2.3}$$

For the statistics $C_\phi \equiv C_{\phi\phi_0}$, Taneichi *et al.* [21] considered the following approximation for the distribution of $C_\phi$ under $H_0^g$.

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

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 Taneichi *et al.* [17] in the case of a multinomial goodness-of-fit test, is a discontinuous term to account for the discontinuity. By using the continuous term $J_1^{g,\phi}(x)$, a transformation for $C_\phi$ that improves the speed of convergence to a $\chi^2$ limiting distribution is constructed. Let $J_1^{g,\phi\phi^*}(x)$ be a continuous term of the approximation of $\Pr\{C_{\phi\phi^*} \le x|H_0^g\}$. Similarly, in this paper, we construct the transformation for $C_{\phi\phi^*}$ by using $J_1^{g,\phi\phi^*}(x)$. With regard to evaluation of the $J_1^{g,\phi\phi^*}(x)$ term, we obtain the following theorem.

**Theorem 2.1.** *When $g^{-1}$ and $\phi^*$ are fourth time continuously differentiable functions and $\phi$ is a fifth time continuously differentiable function, under Assumption 2.1, the $J_1^{g,\phi\phi^*}(x)$ term is evaluated as*

$$J_1^{g,\phi\phi^*}(x) = \Pr\{\chi^2_{N-p} \le x\} + \frac{1}{n}\sum_{j=0}^{3} v_j^{g,\phi\phi^*} \Pr\{\chi^2_{N-p+2j} \le x\} + O(n^{-2}), \tag{2.4}$$

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

$$v_0^{g,\phi\phi^*} = \frac{1}{24}(-\Gamma_4),$$

$$v_1^{g,\phi\phi^*} = \frac{1}{24}\Big[\Gamma_1\phi^{(4)}(1) + \Gamma_2\{\phi'''(1)+1\}^2 + (2\Gamma_1+\Gamma_3)\phi'''(1) + (\Gamma_3+\Gamma_4) + \Delta\Big],$$

$$v_2^{g,\phi\phi^*} = \frac{1}{24}\left[-\Gamma_1\phi^{(4)}(1) - 2\Gamma_2\{\phi'''(1)+1\}^2 - (2\Gamma_1+\Gamma_3)\phi'''(1) - \Gamma_3 - \Delta\right],$$

$$v_3^{g,\phi\phi^*} = \frac{1}{24}\Gamma_2\{\phi'''(1)+1\}^2,$$

*where*

$$\Delta = \Gamma_5\{\phi^{*'''}(1)+1\}\{\phi^{*'''}(1) - 2\phi'''(1) - 1\},$$

$$\Gamma_1 = -3(A_1 - 2A_3 + A_6), \quad \Gamma_2 = 5A_2 - 12A_4 + 9A_7 - 3B_1 + 6B_2 - 2B_4 - 3B_7,$$

$$\Gamma_3 = 2(3A_1 - 2A_2 - 6A_3 + 6A_4 + 3A_5 + 3A_6 - 6A_7 - 3A_8 - 3B_3 + 2B_4 + 3B_8),$$

$$\Gamma_4 = 6A_1 - 4A_2 - 6A_6 + 12A_8 - 3A_9 + 4B_4 - 12B_5 + 6B_6 - 3B_9,$$

$$\Gamma_5 = -3(2A_4 - 4A_7 + B_1 - 2B_2 + 2B_4 + B_7),$$

$$A_1 = \sum_{\alpha=1}^{N} \frac{1 - 3\pi_\alpha^g + 3(\pi_\alpha^g)^2}{\mu_\alpha \pi_\alpha^g (1 - \pi_\alpha^g)}, \quad A_2 = \sum_{\alpha=1}^{N} \frac{(1 - 2\pi_\alpha^g)^2}{\mu_\alpha \pi_\alpha^g (1 - \pi_\alpha^g)},$$

$$A_3 = \sum_{\alpha=1}^{N} \frac{1 - 3\pi_\alpha^g + 3(\pi_\alpha^g)^2}{(\pi_\alpha^g)^2 (1 - \pi_\alpha^g)^2} G_1(\alpha)^2 \sigma_{\alpha\alpha}, \quad A_4 = \sum_{\alpha=1}^{N} \frac{(1 - 2\pi_\alpha^g)^2}{(\pi_\alpha^g)^2 (1 - \pi_\alpha^g)^2} G_1(\alpha)^2 \sigma_{\alpha\alpha},$$

$$A_5 = \sum_{\alpha=1}^{N} \frac{1 - 2\pi_\alpha^g}{\pi_\alpha^g (1 - \pi_\alpha^g)} G_2(\alpha) \sigma_{\alpha\alpha}, \quad A_6 = \sum_{\alpha=1}^{N} \frac{\mu_\alpha (1 - 3\pi_\alpha^g + 3(\pi_\alpha^g)^2)}{(\pi_\alpha^g)^3 (1 - \pi_\alpha^g)^3} G_1(\alpha)^4 \sigma_{\alpha\alpha}^2,$$

$$A_7 = \sum_{\alpha=1}^{N} \frac{\mu_\alpha (1 - 2\pi_\alpha^g)^2}{(\pi_\alpha^g)^3 (1 - \pi_\alpha^g)^3} G_1(\alpha)^4 \sigma_{\alpha\alpha}^2,$$

$$A_8 = \sum_{\alpha=1}^{N} \frac{\mu_\alpha (1 - 2\pi_\alpha^g)}{(\pi_\alpha^g)^2 (1 - \pi_\alpha^g)^2} G_1(\alpha)^2 G_2(\alpha) \sigma_{\alpha\alpha}^2, \quad A_9 = \sum_{\alpha=1}^{N} \frac{\mu_\alpha}{\pi_\alpha^g (1 - \pi_\alpha^g)} G_2(\alpha)^2 \sigma_{\alpha\alpha}^2,$$

$$B_1 = \sum_{\alpha=1}^{N} \sum_{\gamma=1}^{N} \frac{1 - 2\pi_\alpha^g}{\pi_\alpha^g (1 - \pi_\alpha^g)} \frac{1 - 2\pi_\gamma^g}{\pi_\gamma^g (1 - \pi_\gamma^g)} G_1(\alpha) G_1(\gamma) \sigma_{\alpha\gamma},$$

$$B_2 = \sum_{\alpha=1}^{N} \sum_{\gamma=1}^{N} \frac{\mu_\alpha (1 - 2\pi_\alpha^g)}{(\pi_\alpha^g)^2 (1 - \pi_\alpha^g)^2} \frac{1 - 2\pi_\gamma^g}{\pi_\gamma^g (1 - \pi_\gamma^g)} G_1(\alpha)^3 G_1(\gamma) \sigma_{\alpha\alpha} \sigma_{\alpha\gamma},$$

$$B_3 = \sum_{\alpha=1}^{N} \sum_{\gamma=1}^{N} \frac{\mu_\alpha}{\pi_\alpha^g (1 - \pi_\alpha^g)} \frac{1 - 2\pi_\gamma^g}{\pi_\gamma^g (1 - \pi_\gamma^g)} G_1(\alpha) G_2(\alpha) G_1(\gamma) \sigma_{\alpha\alpha} \sigma_{\alpha\gamma},$$

$$B_4 = \sum_{\alpha=1}^{N} \sum_{\gamma=1}^{N} \frac{\mu_\alpha (1 - 2\pi_\alpha^g)}{(\pi_\alpha^g)^2 (1 - \pi_\alpha^g)^2} \frac{\mu_\gamma (1 - 2\pi_\gamma^g)}{(\pi_\gamma^g)^2 (1 - \pi_\gamma^g)^2} G_1(\alpha)^3 G_1(\gamma)^3 \sigma_{\alpha\gamma}^3,$$

$$B_5 = \sum_{\alpha=1}^{N} \sum_{\gamma=1}^{N} \frac{\mu_\alpha}{\pi_\alpha^g (1 - \pi_\alpha^g)} \frac{\mu_\gamma (1 - 2\pi_\gamma^g)}{(\pi_\gamma^g)^2 (1 - \pi_\gamma^g)^2} G_1(\alpha) G_2(\alpha) G_1(\gamma)^3 \sigma_{\alpha\gamma}^3,$$

$$B_6 = \sum_{\alpha=1}^{N} \sum_{\gamma=1}^{N} \frac{\mu_\alpha}{\pi_\alpha^g (1 - \pi_\alpha^g)} \frac{\mu_\gamma}{\pi_\gamma^g (1 - \pi_\gamma^g)} G_1(\alpha) G_2(\alpha) G_1(\gamma) G_2(\gamma) \sigma_{\alpha\gamma}^3,$$

$$B_7 = \sum_{\alpha=1}^{N} \sum_{\gamma=1}^{N} \frac{\mu_\alpha (1 - 2\pi_\alpha^g)}{(\pi_\alpha^g)^2 (1 - \pi_\alpha^g)^2} \frac{\mu_\gamma (1 - 2\pi_\gamma^g)}{(\pi_\gamma^g)^2 (1 - \pi_\gamma^g)^2} G_1(\alpha)^3 G_1(\gamma)^3 \sigma_{\alpha\alpha} \sigma_{\alpha\gamma} \sigma_{\gamma\gamma},$$

$$B_8 = \sum_{\alpha=1}^{N}\sum_{\gamma=1}^{N} \frac{\mu_\alpha}{\pi_\alpha^g(1-\pi_\alpha^g)} \frac{\mu_\gamma(1-2\pi_\gamma^g)}{(\pi_\gamma^g)^2(1-\pi_\gamma^g)^2} G_1(\alpha)G_2(\alpha)G_1(\gamma)^3\sigma_{\alpha\alpha}\sigma_{\alpha\gamma}\sigma_{\gamma\gamma},$$

$$B_9 = \sum_{\alpha=1}^{N}\sum_{\gamma=1}^{N} \frac{\mu_\alpha}{\pi_\alpha^g(1-\pi_\alpha^g)} \frac{\mu_\gamma}{\pi_\gamma^g(1-\pi_\gamma^g)} G_1(\alpha)G_2(\alpha)G_1(\gamma)G_2(\gamma)\sigma_{\alpha\alpha}\sigma_{\alpha\gamma}\sigma_{\gamma\gamma},$$

$$G_i(\alpha) = u^{(i)}(\boldsymbol{x}'_\alpha\boldsymbol{\beta}) \quad (\alpha = 1,\dots,N,\ i = 1,2),$$

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

$$\sigma_{\alpha\gamma} = \sum_{l=1}^{p}\sum_{m=1}^{p} \kappa^{l,m} x_{\alpha l} x_{\gamma m} \qquad (\alpha,\gamma = 1,\dots,N),$$

$$\kappa_{l,m} = \sum_{\lambda=1}^{N} \mu_\lambda\{\pi_\lambda^g(1-\pi_\lambda^g)\}^{-1} G_1(\lambda)^2 x_{\lambda l} x_{\lambda m} \qquad (l,m = 1,\dots,p),$$

*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 2.1 is shown in Appendix. From Theorem 2.1, we can verify the following. The coefficients $v_j^{g,\phi\phi^*}$ $(j = 0,1,2,3)$ satisfy the relation $\sum_{j=0}^{3} v_j^{g,\phi\phi^*} = 0$. The coefficients $v_0^{g,\phi\phi^*}$ and $v_3^{g,\phi\phi^*}$ are not dependent on $\phi^*$. When $\phi^* = \phi_0$, coefficients coincide with those for the family of statistics $C_\phi$ shown in Theorem 1 of Taneichi *et al.* [21].

If we apply $\phi_a$ as $\phi$ and $\phi_b$ as $\phi^*$ in Theorem 2.1, we obtain the following corollary for the statistic $R^{a,b}$ based on power divergence.

**Corollary 2.1.** *When the statistic is* $R^{a,b}$ *given by* (1.5) *and* $g^{-1}$ *is a fourth time continuously differentiable function, under Assumption 2.1, the* $J_1^{g,\phi\phi^*}(x)$ *term is evaluated as*

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

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

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

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

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

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

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

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

$$T_B = \left(1 + \frac{2a_0}{fn}\right) T. \tag{3.1}$$

Then, it holds that

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

This result is known as a Bartlett adjustment. Lawley [8], 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 $T_I$ defined by

$$\begin{aligned} T_I = (n\alpha + \beta)^2 \log\Bigg[1 + \frac{1}{(n\alpha)^2}\Bigg\{T + \frac{1}{n\alpha}(T^2 + \gamma T^3) \\ + \frac{1}{(n\alpha)^2}\left(\frac{1}{3}T^3 + \frac{3\gamma}{4}T^4 + \frac{9\gamma^2}{20}T^5\right)\Bigg\}\Bigg], \end{aligned} \tag{3.2}$$

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

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

The proof of the results for transformation of $T_I$ is given by Yanagihara [22]. The proof is derived by applying the idea of Kakizawa [6] to the theory of improved transformation given by Fujikoshi [5].

Applying the evaluation (2.4) given by Theorem 2.1 to the above transformed statistics $T_B$ given by (3.1) and $T_I$ given by (3.2), we construct transformations for improving small-sample accuracy of the $\chi^2$ approximation of the distribution of $C_\phi$ under $H_0^g$.

When $\phi$ and $\phi^*$ satisfy

$$\phi'''(1) = -1, \quad \phi^{(4)}(1) = 2 \quad \text{and} \quad {\phi^*}'''(1) = -1, \tag{3.3}$$

equations $v_1^{g,\phi\phi^*} = -v_0^{g,\phi\phi^*}$ and $v_2^{g,\phi\phi^*} = v_3^{g,\phi\phi^*} = 0$ hold in Theorem 2.1. Then, we can consider Bartlett-type adjustment

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

On the other hand, when $\phi$ does not satisfy (3.3), we can consider the transformed statistic

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

where

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

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

Practically, we may use estimate $\hat{v}_j^{g,\phi\phi^*}$ $(j = 0, 2, 3)$ obtained by substituting minimum $\phi^*$-divergence estimate $\hat{\boldsymbol{\beta}}^{g\phi^*}$ for true value $\boldsymbol{\beta}$ in $v_j^{g,\phi\phi^*}$ $(j = 0, 2, 3)$. Therefore, when $\phi$ and $\phi^*$ satisfy (3.3), we propose the statistic $\tilde{C}^B_{\phi\phi^*}$ that is obtained by substituting $\hat{v}_0^{g,\phi\phi^*}$ for $v_0^{g,\phi\phi^*}$ in $C^B_{\phi\phi^*}$, that is,

$$\tilde{C}^B_{\phi\phi^*} = \left\{1 + \frac{2\hat{v}_0^{g,\phi\phi^*}}{n(N-p)}\right\} C_{\phi\phi^*}. \tag{3.4}$$

Similarly, when $\phi$ and $\phi^*$ do not satisfy (3.3), we also propose the statistic $\tilde{C}^I_{\phi\phi^*}$ that is obtained by substituting $\hat{v}_j^{g,\phi\phi^*}$ $(j = 0, 2, 3)$ for $v_j^{g,\phi\phi^*}$ $(j = 0, 2, 3)$ in $C^I_{\phi\phi^*}$.

In the case of power divergence statistic $R^{a,b} = C_{\phi_a\phi_b}$ using the minimum power divergence estimator, condition (3.3) is satisfied if and only if $a = 0$

and $b = 0$ (log likelihood ratio statistic). Then, we consider the transformed statistic given by (3.4) when $a = 0$ and $b = 0$ and put $\tilde{R}_B^{0,0} = \tilde{C}_{\phi_0\phi_0}^B$. When the link function $g$ is a logit link function, statistic $\tilde{R}_B^{0,0}$ coincides with the statistic $\tilde{D}$ proposed by (3.4) of Taneichi *et al.* [20]. On the other hand, we consider statistic $\tilde{C}_{\phi_a\phi_b}^I$ when $a \neq 0$ or $b \neq 0$ and put $\tilde{R}_I^{a,b} = \tilde{C}_{\phi_a\phi_b}^I$ ($a \neq 0$ or $b \neq 0$).

We summarize the difference and relation between $T_B$ and $T_I$. Transformed statistic $T_B$ is a simple monotone transformation of $C_\phi$ constructed by a linear function whose intercept is zero. On the other hand, transformed statistic $T_I$ is a monotone transformation of $C_\phi$ constructed by logarithm of quintic function. It is much more complicated than $T_B$. Then, from point of view of stability, $T_I$ seems to be inferior to $T_B$. However, for Cressie and Read family of statistics $R^{a,b}$, $T_B$ increases the speed of convergence to chi-square distribution only for the statistic in the case of $a = b = 0$, that is, the log-likelihood ratio statistic. Therefore, for improving the other statistics, statistic $T_I$ is developed.

## §4. Performance of transformed statistics

In this section, we compare the performance of transformed statistics $\tilde{R}_I^{a,b}$ ($a \neq 0$ or $b \neq 0$) with that of the original power divergence statistics $R^{a,b}$ using the minimum power divergence estimator by the Monte Carlo procedure. The performance of transformed statistic $\tilde{R}_B^{0,0}$ for complementary log-log link $g_0$ and probit link $g_P$ is shown in Fig.1, Fig.2 and Fig.3 of Taneichi *et al.* [21]. We consider a generalized linear model given by (1.1) with $p = 2$ and $x_{\alpha 1} = 1$ and $x_{\alpha 2} = x_\alpha$ ($\alpha = 1, \ldots, N$).

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

$$\pi_\alpha^{g*} = g^{-1}(\beta_1^* + \beta_2^* x_\alpha) \quad (\alpha = 1, \ldots, N). \tag{4.1}$$

As a link function $g$, we consider the family of link functions given by Aranda-Ordaz [1],

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

that depend on parameter $c$. $g_c$ include the logit link $g_1$ and complementary log-log link $g_0$ as a limit. We also consider the probit link $g_P(t) = \Phi^{-1}(t)$, where $\Phi$ is the cumulative distribution function of a standard normal distribution.

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 $\alpha$, we generate $n_\alpha$ $(\alpha = 1, \ldots, N)$ binomial random numbers that are distributed according to $B(1, \pi_\alpha^{g*})$ $(\alpha = 1, \ldots, N)$. From them, we calculate the number of successes $Y_\alpha$ $(\alpha = 1, \ldots, N)$ and the minimum $\phi_b$-divergence estimates $\hat{\beta}_1^{g\phi_b}$ and $\hat{\beta}_2^{g\phi_b}$ for the parameters $\beta_1$ and $\beta_2$. Using the estimates, we calculate the values $\pi_\alpha(\hat{\boldsymbol{\beta}}^{g\phi_b})$ $(\alpha = 1, \ldots, N)$, where $\hat{\boldsymbol{\beta}}^{g\phi_b} = (\hat{\beta}_1^{g\phi_b}, \hat{\beta}_2^{g\phi_b})'$, and observed values of the statistics $R^{a,b}$, $\tilde{R}_I^{a,b}$ ($a \neq 0$ or $b \neq 0$). 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, $\chi^2_{N-p}(\varepsilon)$. 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)

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

(II)

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

(III)

$$\boldsymbol{X} = \left( \begin{matrix} 1 & 1 & 1 & 1 \\ \log(2.7) & \log(3.0) & \log(3.3) & \log(3.6) \end{matrix} \right.$$
$$\left. \begin{matrix} 1 & 1 & 1 & 1 \\ \log(3.9) & \log(4.2) & \log(4.5) & \log(4.8) \end{matrix} \right)'.$$

(IV)

$$\boldsymbol{X} = \left( \begin{matrix} 1 & 1 & 1 & 1 \\ \log(2.85) & \log(3.05) & \log(3.25) & \log(3.45) \end{matrix} \right.$$
$$\left. \begin{matrix} 1 & 1 & 1 & 1 \\ \log(3.65) & \log(3.85) & \log(4.05) & \log(4.25) \end{matrix} \right)'.$$

For each case, we consider a sample design $n_1 = \cdots = n_8 = n_*$.

We investigate the performance for all combinations of the two true parameters (i) and (ii), four design matrices (I), (II), (III) and (IV), and the sample design with $n_* = 20$. Some of the results of the investigations are shown in figures as follows.

Fig.1 shows the absolute values of index $I$ when the test statistic is $R^{0.2,0.2}$ and models are given by link functions $g_0$ (complementary log-log model), $g_{1/2}$, $g_1$ (logistic regression model) and $g_P$ (probit model) in the case of true parameters (i) and (ii), design matrices (I)–(IV), and significance level $\varepsilon =$ 0.01, 0.05 and 0.10. Fig.2 and Fig.3 show the absolute values of index $I$ when the test statistics are $R^{0.0,1.0}$ and $R^{1.0,1.0}$ in the same models and situations as those in the explanation of Fig.1, respectively.

From Fig.1 and Fig.2, we find that the performance of transformed statistics $\tilde{R}_I^{0.2,0.2}$ and $\tilde{R}_I^{0.0,1.0}$ is better than that of original statistics $R^{0.2,0.2}$ and $R^{0.0,1.0}$, respectively, when the models are given by the link functions $g_0$ (complementary log-log model), $g_{1/2}$, $g_1$ (logistic regression model) and $g_P$ (probit model) for the two true parameters, all design matrix cases, and sample design $n_* = 20$. From Fig.3, we find that the performance of transformed statistic $\tilde{R}_I^{1.0,1.0}$ is better than that of original statistic $R^{1.0,1.0}$ when the true parameter is type (i). However, when the true parameter is type (ii), the performance of the transformed statistic is not better than that of the original statistic.

Consequently, from Figs.1–3 and other simulation results, we conclude as follows. The performance of $\tilde{R}_I^{a,b}$ $(0 < a \leq 1, 0 < b \leq 1)$ is usually better than that of original statistic $R^{a,b}$ $(0 < a \leq 1, 0 < b \leq 1)$ when the models are given by the link functions $g_0$ (complementary log-log model), $g_{1/2}$, $g_1$ (logistic regression model) and $g_P$ (probit model) under the conditions of the simulation. However, as shown in Fig.3, when the chi-square approximation of the original statistic performs very well, approximation of the transformed statistic sometimes does not perform better than the original statistic. That is, when the chi-square approximation of the original statistic already performs very well, the transformed statistic sometimes cannot improve the performance of chi-square approximation.

Next, we compare the power of transformed statistics $\tilde{R}_I^{a,b}$ ($a \neq 0$ or $b \neq 0$) with that of the original statistics $R^{a,b}$. The power of transformed statistic $\tilde{R}_B^{0.0,0.0}$ for complementary log-log link $g_0$ and probit link $g_P$ is shown in Fig.6 and Fig.7 of Taneichi *et al.* [21]. Against the null model given by (4.1), we consider an alternative model:

$$H_1^g : \pi_\alpha^{g*} = g^{-1}(\beta_1^* + \beta_2^* x_\alpha) + \delta_\alpha \quad (\alpha = 1, \ldots, 8), \tag{4.2}$$

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 (4.2) by using simulated exact critical values of statistics. We investigate the average power for all combinations of the two true parameters (i) and (ii), four design matrices (I)–(IV), and sample design $n_* = 20$. In the investigation, the number of repetitions is $D = 10^6$. Some of the results of the investigations are shown in figures as follows. Figs.4, 5 and 6 show the power of statistics corresponding to the cases in Figs.1, 2 and 3, respectively.

From Figs.4–6 and other simulation results, we conclude that the power against $H_1^g$ given by (16) of the transformed statistics $\tilde{R}_I^{a,b}$ $(0 < a \leq 1, 0 < b \leq 1)$ is not so different from that of the original power divergence statistic $R^{a,b}$ in the models based on link functions $g_c$ $(c = 0, 1/2, 1)$ and $g_P$.

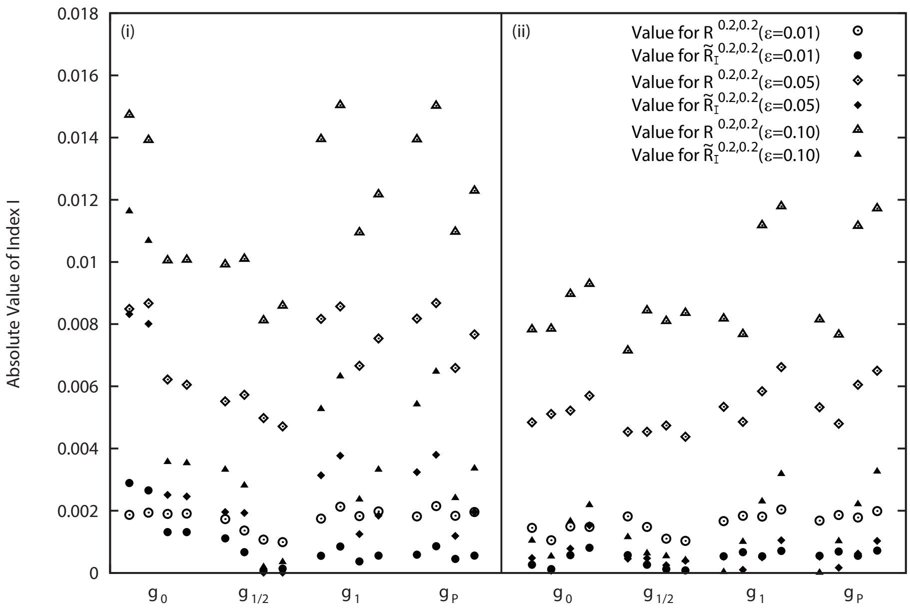


Figure 1: Absolute value of index $I$ when the original test statistic is $R^{0.2,0.2}$ and models are given by link functions $g_0$, $g_{1/2}$, $g_1$ and $g_P$ for true parameters (i) and (ii) and sample design $n_*$ =20: ○, ◇ and △ are the values for $R^{0.2,0.2}$ when $\varepsilon$ =0.01, 0.05 and 0.10, respectively, and ●,◆ and ▲ are the values for $\tilde{R}_I^{0.2,0.2}$ 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).

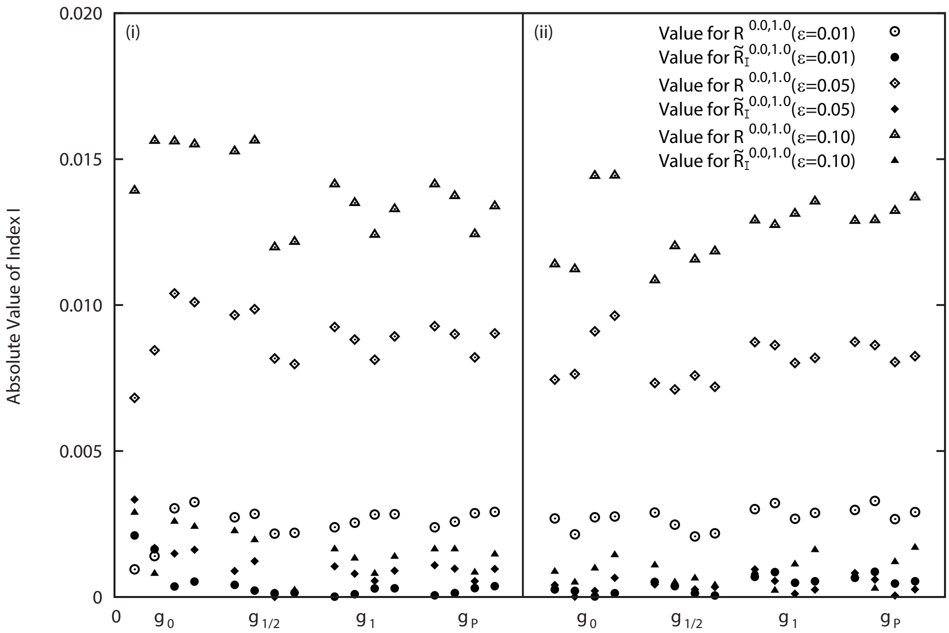


Figure 2: Absolute value of index $I$ when the original test statistic is $R^{0.0,1.0}$.

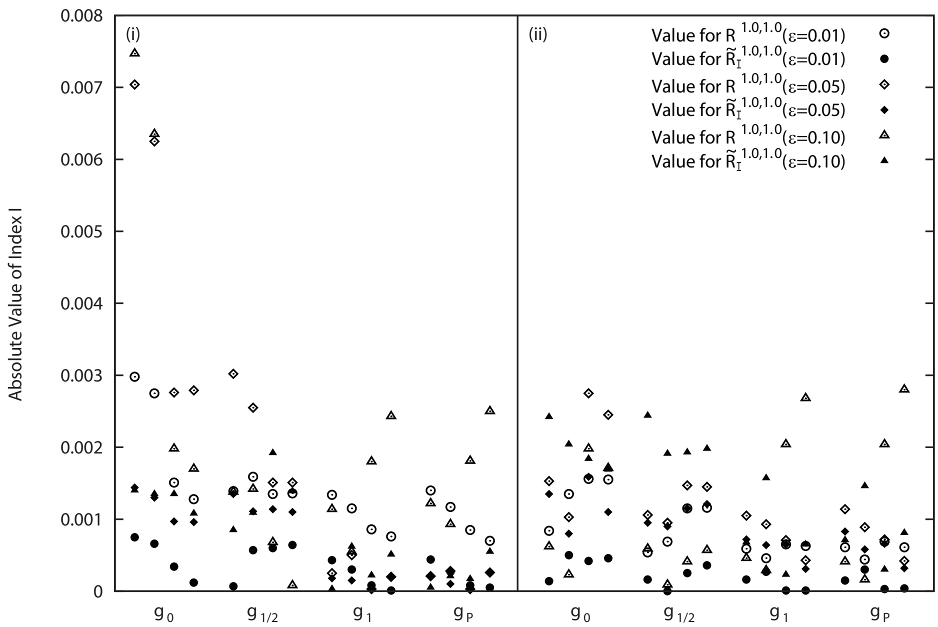


Figure 3: Absolute value of index $I$ when the original test statistic is $R^{1.0,1.0}$.

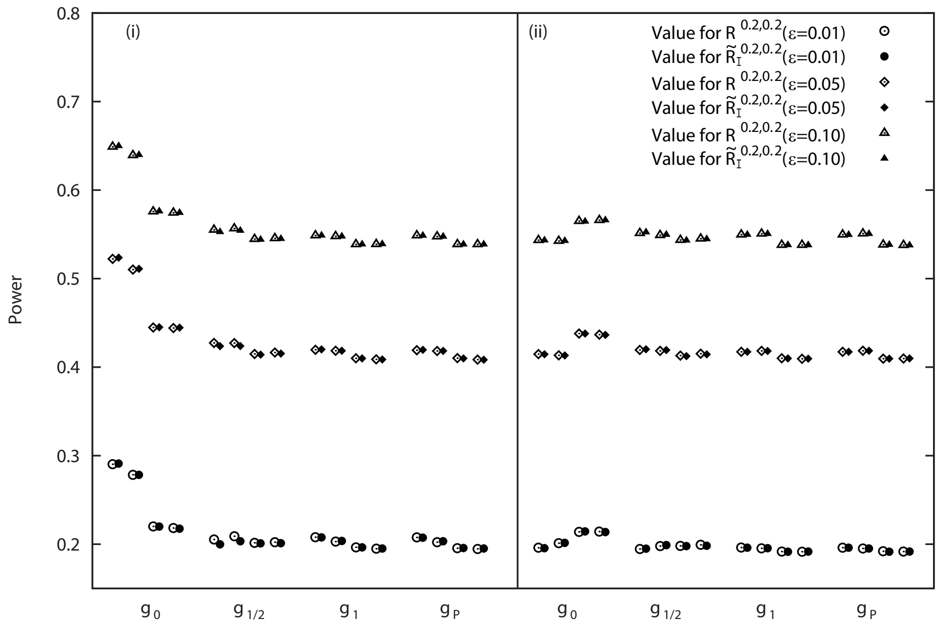


Figure 4: Simulated average power $P$ against an alternative model (4.2) when the original test statistic is $R^{0.2,0.2}$ and models are given by link functions $g_0$, $g_{1/2}$, $g_1$ and $g_P$ for true parameters (i) and (ii) and sample design $n_* = 20$: $\circ$, $\Diamond$ and $\triangle$ are the values for $R^{0.2,0.2}$ when $\varepsilon$ =0.01, 0.05 and 0.10, respectively, and $\bullet$,$\blacklozenge$ and $\blacktriangle$ are the values for $\tilde{R}_I^{0.2,0.2}$ 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).

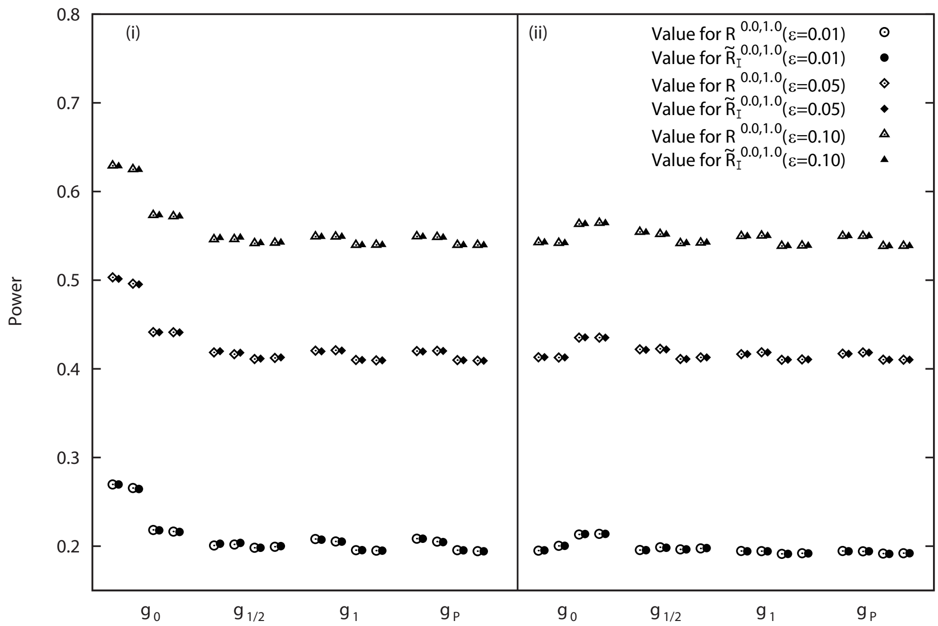


Figure 5: Simulated average power $P$ against an alternative model (4.2) when the original test statistic is $R^{0.0,1.0}$.

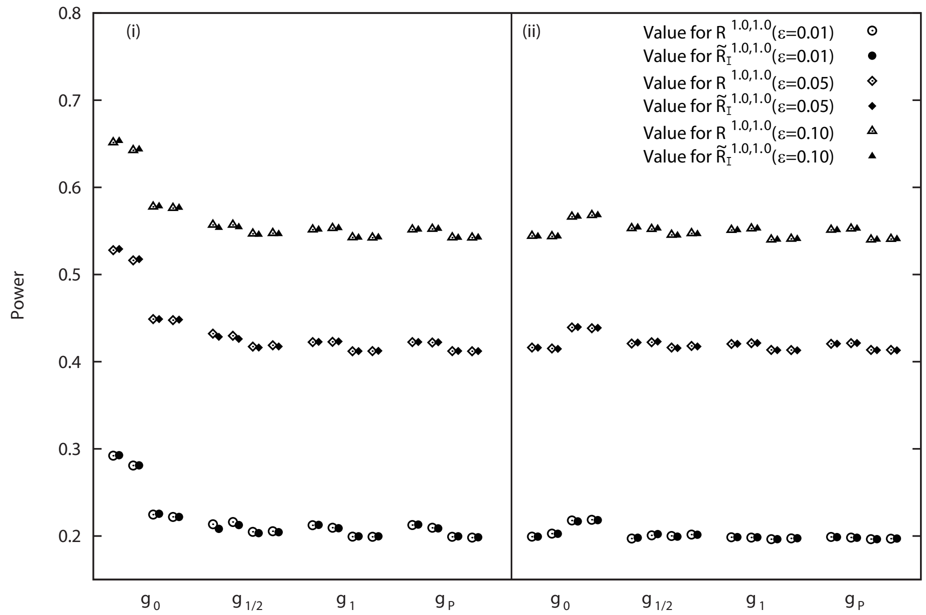


Figure 6: Simulated average power $P$ against an alternative model (4.2) when the original test statistic is $R^{1.0,1.0}$.

## §5. Appendix: Proof of Theorem 2.1

By transformation (2.1), statistic $C_{\phi\phi^*}$ can be rewritten as

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

If we regard

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

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

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

as the distribution function of $C_{\phi\phi^*}(\boldsymbol{W})$, where

$$U^g_{\phi\phi^*}(x) = \{\boldsymbol{w} = (w_1, \ldots, w_N)' : C_{\phi\phi^*}(\boldsymbol{w}) \leq x\}.$$

So, the characteristic function of $C_{\phi\phi^*}(\boldsymbol{W})$ is calculated as

$$\text{(A1)} \quad \psi^g_{\phi\phi^*}(u) = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} [\exp\{iuC_{\phi\phi^*}(\boldsymbol{w})\}]\, h^g(\boldsymbol{w}) \times \left\{1 + \frac{1}{\sqrt{n}} h^g_1(\boldsymbol{w}) + \frac{1}{n} h^g_2(\boldsymbol{w}) + \frac{1}{n\sqrt{n}} h^g_3(\boldsymbol{w})\right\} d\boldsymbol{w}.$$

We can expand $C_{\phi\phi^*}(\boldsymbol{w})$ as

$$\text{(A2)} \quad C_{\phi\phi^*}(\boldsymbol{w}) = \tau^g_0(\boldsymbol{w}) + \frac{1}{\sqrt{n}} \tau^{g,\phi}_1(\boldsymbol{w}) + \frac{1}{n} \tau^{g,\phi\phi^*}_2(\boldsymbol{w}) + \frac{1}{n\sqrt{n}} \tau^{g,\phi\phi^*}_3(\boldsymbol{w}) + O(n^{-2}),$$

where

$$\tau^g_0(\boldsymbol{w}) = \boldsymbol{w}'(\Omega^{-1} - \Xi)\boldsymbol{w},$$

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

$$\xi_{\alpha\beta} = \frac{\sqrt{\mu_\alpha} G_1(\alpha)}{\pi^g_\alpha(1 - \pi^g_\alpha)} \frac{\sqrt{\mu_\beta} G_1(\beta)}{\pi^g_\beta(1 - \pi^g_\beta)} \sigma_{\alpha\beta} \quad (\alpha, \beta = 1, \ldots, N),$$

$$\tau^{g,\phi}_1(\boldsymbol{w}) = \sum_{a=0}^{3} \left( \sum_{\alpha=1}^{N} B^1_{a+1}(\alpha) C_{1(\alpha)}(\boldsymbol{w})^{3-a} w^a_\alpha \right),$$

$$\begin{aligned} \tau^{g,\phi\phi^*}_2(\boldsymbol{w}) &= \sum_{a=0}^{2} \left( \sum_{\alpha=1}^{N} B^2_{a+1}(\alpha) C^{\phi^*}_{2(\alpha)}(\boldsymbol{w})^{2-a} C_{1(\alpha)}(\boldsymbol{w})^{2a} \right) \\ &\quad + \sum_{a=0}^{1} \left( \sum_{\alpha=1}^{N} B^2_{a+4}(\alpha) C^{\phi^*}_{2(\alpha)}(\boldsymbol{w})^{1-a} C_{1(\alpha)}(\boldsymbol{w})^{1+2a} w_\alpha \right) \\ &\quad + \sum_{a=0}^{1} \left( \sum_{\alpha=1}^{N} B^2_{a+6}(\alpha) C^{\phi^*}_{2(\alpha)}(\boldsymbol{w})^{1-a} C_{1(\alpha)}(\boldsymbol{w})^{2a} w^2_\alpha \right) \\ &\quad + \sum_{\alpha=1}^{N} B^2_8(\alpha) C_{1(\alpha)}(\boldsymbol{w}) w^3_\alpha + \sum_{\alpha=1}^{N} B^2_9(\alpha) w^4_\alpha, \end{aligned}$$

$$\begin{aligned} B^1_1(\alpha) &= \frac{\mu_\alpha}{3(\pi^g_\alpha)^2(1 - \pi^g_\alpha)^2} \big\{3\pi^g_\alpha(1 - \pi^g_\alpha) G_1(\alpha) G_2(\alpha) \\ &\qquad - (3 + \phi'''(1))(1 - 2\pi^g_\alpha) G_1(\alpha)^3\big\}, \end{aligned}$$

$$\begin{aligned} B^1_2(\alpha) &= \frac{\sqrt{\mu_\alpha}}{(\pi^g_\alpha)^2(1 - \pi^g_\alpha)^2} \big\{-\pi^g_\alpha(1 - \pi^g_\alpha) G_2(\alpha) \\ &\qquad + (2 + \phi'''(1))(1 - 2\pi^g_\alpha) G_1(\alpha)^2\big\}, \end{aligned}$$

$$B_3^1(\alpha) = -\frac{\left(1+\phi'''(1)\right)(1-2\pi_\alpha^g)G_1(\alpha)}{(\pi_\alpha^g)^2(1-\pi_\alpha^g)^2}, \quad B_4^1(\alpha) = \frac{\phi'''(1)(1-2\pi_\alpha^g)}{3\sqrt{\mu_\alpha}(\pi_\alpha^g)^2(1-\pi_\alpha^g)^2},$$

$$B_1^2(\alpha) = \frac{\mu_\alpha G_1(\alpha)^2}{\pi_\alpha^g(1-\pi_\alpha^g)}, \quad B_2^2(\alpha) = 3B_1^1(\alpha),$$

$$\begin{aligned} B_3^2(\alpha) &= \frac{\mu_\alpha}{12(\pi_\alpha^g)^3(1-\pi_\alpha^g)^3}\{(\pi_\alpha^g)^2(1-\pi_\alpha^g)^2(3G_2(\alpha)^2+4G_1(\alpha)G_3(\alpha)) \\ &\quad -6(3+\phi'''(1))\pi_\alpha^g(1-\pi_\alpha^g)(1-2\pi_\alpha^g)G_1(\alpha)^2G_2(\alpha) \\ &\quad +(12+8\phi'''(1)+\phi^{(4)}(1))(1-3\pi_\alpha^g+3(\pi_\alpha^g)^2)G_1(\alpha)^4\}, \end{aligned}$$

$$B_4^2(\alpha) = 2B_2^1(\alpha),$$

$$\begin{aligned} B_5^2(\alpha) &= \frac{\sqrt{\mu_\alpha}}{3(\pi_\alpha^g)^3(1-\pi_\alpha^g)^3}\{-(\pi_\alpha^g)^2(1-\pi_\alpha^g)^2G_3(\alpha) \\ &\quad +3(2+\phi'''(1))\pi_\alpha^g(1-\pi_\alpha^g)(1-2\pi_\alpha^g)G_1(\alpha)G_2(\alpha) \\ &\quad -(6+6\phi'''(1)+\phi^{(4)}(1))(1-3\pi_\alpha^g+3(\pi_\alpha^g)^2)G_1(\alpha)^3\}, \end{aligned}$$

$$B_6^2(\alpha) = B_3^1(\alpha),$$

$$\begin{aligned} B_7^2(\alpha) &= \frac{1}{2(\pi_\alpha^g)^3(1-\pi_\alpha^g)^3}\{-(1+\phi'''(1))\pi_\alpha^g(1-\pi_\alpha^g)(1-2\pi_\alpha^g)G_2(\alpha) \\ &\quad +(2+4\phi'''(1)+\phi^{(4)}(1))(1-3\pi_\alpha^g+3(\pi_\alpha^g)^2)G_1(\alpha)^2\}, \end{aligned}$$

$$B_8^2(\alpha) = -\frac{(2\phi'''(1)+\phi^{(4)}(1))(1-3\pi_\alpha^g+3(\pi_\alpha^g)^2)G_1(\alpha)}{3\sqrt{\mu_\alpha}(\pi_\alpha^g)^3(1-\pi_\alpha^g)^3},$$

$$B_9^2(\alpha) = \frac{\phi^{(4)}(1)(1-3\pi_\alpha^g+3(\pi_\alpha^g)^2)}{12\mu_\alpha(\pi_\alpha^g)^3(1-\pi_\alpha^g)^3},$$

$$C_{1(\alpha)}(\boldsymbol{w}) = \sum_{m=1}^p x_{\alpha m}\left(\sum_{k=1}^p \kappa^{m,k}M_k(\boldsymbol{w})\right) \ (\alpha = 1,\ldots,N),$$

$$\begin{aligned} C_{2(\alpha)}^{\phi^*}(\boldsymbol{w}) &= \sum_{m=1}^p x_{\alpha m}\left\{\sum_{k=1}^p M^{m,k}(\boldsymbol{w})M_k(\boldsymbol{w}) + \sum_{k=1}^p \kappa^{m,k}S_k^{\phi^*}(\boldsymbol{w})\right. \\ &\quad \left. +\frac{1}{2}\sum_{k_1=1}^p \cdots \sum_{k_5=1}^p \kappa^{m,k_3}\kappa^{k_1,k_4}\kappa^{k_2,k_5}\kappa_{k_3,k_4,k_5}^{\phi^*}M_{k_1}(\boldsymbol{w})M_{k_2}(\boldsymbol{w})\right\} \\ &\qquad (\alpha = 1,\ldots,N), \end{aligned}$$

$$M_k(\boldsymbol{w}) = \sum_{\lambda=1}^N \sqrt{\mu_\lambda}x_{\lambda k}G_1(\lambda)\{\pi_\lambda^g(1-\pi_\lambda^g)\}^{-1}w_\lambda \quad (k = 1,\ldots,p),$$

$$\begin{aligned} Q_{i,j}^{\phi^*}(\boldsymbol{w}) &= \sum_{\lambda=1}^N \sqrt{\mu_\lambda}x_{\lambda i}x_{\lambda j}\left\{-(2+\phi^{*\prime\prime\prime}(1))\frac{(1-2\pi_\lambda^g)G_1(\lambda)^2}{(\pi_\lambda^g)^2(1-\pi_\lambda^g)^2}\right. \\ &\quad \left. +\frac{G_2(\lambda)}{\pi_\lambda^g(1-\pi_\lambda^g)}\right\}w_\lambda \quad (i,j = 1,\ldots,p), \end{aligned}$$

$$\kappa^{\phi^*}_{i,j,k} = \sum_{\lambda=1}^{N}\mu_\lambda x_{\lambda i}x_{\lambda j}x_{\lambda k}\left\{(3+\phi^{*'''}(1))\frac{(1-2\pi^g_\lambda)G_1(\lambda)^3}{(\pi^g_\lambda)^2(1-\pi^g_\lambda)^2} -3\frac{G_1(\lambda)G_2(\lambda)}{\pi^g_\lambda(1-\pi^g_\lambda)}\right\} \quad (i,j,k=1,\ldots,p),$$

$$S^{\phi^*}_k(\boldsymbol{w}) = \frac{1}{2}\{1+\phi^{*'''}(1)\}\sum_{\lambda=1}^{N}x_{\lambda k}\frac{(1-2\pi^g_\lambda)G_1(\lambda)}{(\pi^g_\lambda)^2(1-\pi^g_\lambda)^2}w_\lambda^2 \quad (k=1,\ldots,p),$$

$$G_i(\alpha) = u^{(i)}(\boldsymbol{x}'_\alpha\boldsymbol{\beta}) \quad (\alpha=1,\ldots,N,\ i=1,2,3),$$

$Q^{\phi^*}(\boldsymbol{w}) = (Q^{\phi^*}_{i,j}(\boldsymbol{w}))$ is a $p\times p$ matrix, $M^{i,j}(\boldsymbol{w})$ is the $(i,j)$-element of matrix $K^{-1}Q^{\phi^*}(\boldsymbol{w})K^{-1}$, $\Omega$ is defined by (2.3), $\sigma_{\alpha\beta}$ and $K^{-1}=(\kappa^{i,j})$ are defined in Theorem 2.1, and $\tau^{g,\phi\phi^*}_3(\boldsymbol{w})$ is a homogeneous polynomial of degree 5 with respect to variables $w_1,\ldots,w_N$. Then, from (2.2), (A1) and (A2), we obtain

(A3)
$$\psi^g_{\phi\phi^*}(u) = (1-2iu)^{-(N-p)/2} \times\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}(2\pi)^{-N/2}|\Lambda|^{-1/2}\left\{\exp\left(-\frac{1}{2}\boldsymbol{w}'\Lambda^{-1}\boldsymbol{w}\right)\right\} \times\left\{1+\frac{1}{\sqrt{n}}D_1(\boldsymbol{w})+\frac{1}{n}D_2(\boldsymbol{w})+\frac{1}{n\sqrt{n}}D_3(\boldsymbol{w})\right\}d\boldsymbol{w} +O(n^{-2}),$$

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

$$D_1(\boldsymbol{w}) = h^g_1(\boldsymbol{w}) + (iu)\tau^{g,\phi}_1(\boldsymbol{w}),$$

$$D_2(\boldsymbol{w}) = h^g_2(\boldsymbol{w}) + (iu)\tau^{g,\phi}_1(\boldsymbol{w})h^g_1(\boldsymbol{w}) + (iu)\tau^{g,\phi\phi^*}_2(\boldsymbol{w}) + \frac{1}{2}(iu)^2\left\{\tau^{g,\phi}_1(\boldsymbol{w})\right\}^2,$$

and degrees of all terms of polynomial $D_3(\boldsymbol{w})$ are odd. Therefore, by carrying out the integration of (A3), the characteristic function $\psi^g_{\phi\phi^*}(u)$ is expanded as

(A4)
$$\psi^g_{\phi\phi^*}(u) = (1-2iu)^{-(N-p)/2}\left[1+\frac{1}{n}\sum_{j=0}^{3}(1-2iu)^{-j}v^{g,\phi\phi^*}_j + O(n^{-2})\right].$$

By inverting (A4), we obtain (2.4). We have completed the proof of Theorem 2.1.

## Acknowledgments

This research is partially supported by the Grants-in-aid for Scientific Research of Japan Society for the Promotion of Science (C) 24540133.

## References

[1] Aranda-Ordaz, F. J. (1981). On two families of transformations to additivity for binary response data, *Biometrika*, **68**, 357–363.

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

[3] Barndorff-Nielsen, O. E. and Hall, P. (1988). On the level-error after Bartlett adjustment of the likelihood ratio statistic, *Biometrika*, **75**, 374–378.

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

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

[6] Kakizawa, Y. (1996). Higher order monotone Bartlett-type adjustment for some multivariate test statistics, *Biometrika*, **83**, 923–927.

[7] Kullback, S. (1959). *Information Theory and statistics*, New York Wiley.

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

[9] Menéndez, M. L., Pardo, J. A., Pardo, L. and Pardo, M. C. (1997). Asymptotic approximations for the distributions of the $(h, \phi)$-divergence goodness-of-fit statistics: application to Renyi's statistic, *Kybernetes*, **26**(4), 442–452.

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

[11] Pardo, L. (2006). *Statistical inference based on divergence measures*, Chapman & Hall/CRC.

[12] Pardo, J. A. and Pardo, M. C. (2008). Minimum $\phi$-divergence estimator and $\phi$-divergence statistics in generalized linear models with binary data, *Methodol. Comput. Appl. Probab.*, **10**, 357–379.

[13] Read, T. R. C. (1984). Closer asymptotic approximations for the distributions of the power divergence goodness-of-fit statistics, *Ann. Inst. Statist. Math.*, **36**, 59–69.

[14] Read, T. R. C. and Cressie, N. A. C. (1988). *Goodness-of-fit statistics for discrete multivariate data*, Springer.

[15] Sekiya, Y. and Taneichi, N. (2004). Improvement of approximations for the distributions of multinomial goodness-of-fit statistics under nonlocal alternatives, *J. Multivariate Anal.*, **91**, 199–223.

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

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

[18] Taneichi, N., Sekiya, Y. and Suzukawa, A. (2002). Asymptotic approximations for the distributions of the multinomial goodness-of-fit statistics under local alternatives, *J. Multivariate Anal.*, **81**, 335–359.

[19] Taneichi, N. and Sekiya, Y. (2007). Improved transformed statistics for the test of independence in $r \times s$ contingency tables, *J. Multivariate Anal.*, **98**, 1630–1657.

[20] Taneichi, N., Sekiya, Y. and Toyama, J. (2011). Improved transformed deviance statistic for testing a logistic regression model, *J. Multivariate Anal.*, **102**, 1263–1279.

[21] Taneichi, N., Sekiya, Y. and Toyama, J. (2014). Transformed goodness-of-fit statistics for a generalized linear model of binary data, *J. Multivariate Anal.*, **123**, 311–329.

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

[23] Yarnold, J. K. (1972). Asymptotic approximations for the probability that a sum of lattice random vectors lies in a convex set, *Ann. Math. Statist.*, **43**, 1566–1580.

Nobuhiro Taneichi
Department of 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
Kushiro Campus, Hokkaido University of Education
Kushiro 085-8580, Japan
*E-mail*: `sekiya.yuri@k.hokkyodai.ac.jp`

Jun Toyama
The Institute for the Practical Application of Mathematics
Sapporo 063-0001, Japan
*E-mail*: `mandheling@nifty.com`