HOKKAIDO UNIVERSITY

Title	Improved transformed deviance statistic for testing a logistic regression model
Author(s)	Taneichi, Nobuhiro; Sekiya, Yuri; Toyama, Jun
Citation	Journal of Multivariate Analysis, 102(9), 1263-1279 https://doi.org/10.1016/j.jmva.2011.04.010
Issue Date	2011-10
Doc URL	http://hdl.handle.net/2115/47351
Type	article (author version)
File Information	JMA102-9_1263-1279.pdf

Instructions for use

Hokkaido University Collection of Scholarly and Academic Papers : HUSCAP

*Manuscript
Click here to view linked References

Improved transformed deviance statistic for testing a logistic regression model

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

and

Jun Toyama

Division of Computer Science, Graduate School of Information Science and Technology, Hokkaido University, Sapporo 060-0814, Japan

E-mail: jun@main.ist.hokudai.ac.jp

Running head: IMPROVED DEVIANCE FOR LOGISTIC REGRESSION

Address for manuscript correspondence:

Nobuhiro Taneichi

Department of Mathematics and Computer Science,

Graduate School of Science and Engineering,

Kagoshima University,

1-21-35 Korimoto, Kagoshima 890-0065, Japan

Telephone: +81-99-285-8986

Fax: +81-99-285-8986

E-mail: taneichi@sci.kagoshima-u.ac.jp

Abstract

In logistic regression models, we consider the deviance statistic (the log likelihood ratio statistic) D as a goodness-of-fit test statistic. In this paper, we show the derivation of an expression of asymptotic expansion for the distribution of D under a null hypothesis. Using the continuous term of the expression, we obtain Bartlett-type transformed statistic $\tilde{D}$ that improves the speed of convergence to the chi-square limiting distribution of D. By numerical comparison, we find that the transformed statistic $\tilde{D}$ performs much better than D. We also give a real data example of $\tilde{D}$ being more reliable than D for testing a hypothesis.

AMS 2000 subject classifications: 62E20, 62H10.

Keywords: Bartlett adjustment; Deviance; Edgeworth expansion; Logistic regression.

1. Introduction

We consider generalized linear models (Nelder and Wedderburn [9]) in which the response variables are measured on a binary scale. Let N independent random variables $Y_\alpha,\ \alpha = 1, \ldots, N$ corresponding to the number of successes in N different subgroups be distributed according to a binomial distribution $B(n_\alpha, \pi_\alpha),\ \alpha = 1, \ldots, N$. If we use the logit function

$$\text{logit } u \equiv \log\left(\frac{u}{1-u}\right),$$

which is a canonical link function, as a link function, we obtain the following general logistic regression model (general logit model):

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

where $\boldsymbol{x}_\alpha = (x_{\alpha 1}, \ldots, x_{\alpha p})',\ (\alpha = 1, \ldots, N),\ (p < N)$ are covariate vectors and $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_p)'$ is a unknown parameter vector. Let the maximum likelihood estimator of $\boldsymbol{\beta}$ be $\hat{\boldsymbol{\beta}} = (\hat{\beta}_1, \ldots, \hat{\beta}_p)'$, and put $\hat{\pi}_\alpha = \pi_\alpha(\hat{\boldsymbol{\beta}}),\ (\alpha = 1, \ldots, N)$. Here, we consider the deviance statistic (log likelihood ratio statistic)

$$D = 2\sum_{\alpha=1}^{N} n_\alpha \left\{ \frac{Y_\alpha}{n_\alpha} \log\left(\frac{Y_\alpha}{n_\alpha \hat{\pi}_\alpha}\right) + \left(1 - \frac{Y_\alpha}{n_\alpha}\right) \log\left(\frac{1 - \dfrac{Y_\alpha}{n_\alpha}}{1 - \hat{\pi}_\alpha}\right) \right\}. \tag{1.2}$$

Under the null hypothesis

$$H_0 : \text{Model given by (1.1) is correct,} \tag{1.3}$$

it is known that deviance statistic D has a χ^2_{N-p} limiting distribution assuming the condition that

$$n_\alpha / n \to \mu_\alpha \ (0 < \mu_\alpha < 1) \text{ for each } \alpha, \quad \text{as} \quad n \to \infty, \tag{1.4}$$

where $n = \sum_{\alpha=1}^{N} n_\alpha$ and $\sum_{\alpha=1}^{N} \mu_\alpha = 1$. Usually, using large sample results, we use D for a goodness-of-fit test statistic of the logistic regression model.

However, in the case in which all $n_\alpha, (\alpha = 1, \ldots, N)$ are not large enough, approximation by a χ^2_{N-p} limiting distribution to the distribution of D under H_0 become poor. In such a case, there are risks that the hypothesis test based on large sample theory will give results opposite to those of an exact test. In this paper, in order to reduce the risks, we propose a new transformed statistic $\tilde{D}$ of D whose speed of convergence to a chi-square distribution is quicker than D. To construct $\tilde{D}$, we use the following procedure. First, we obtain the asymptotic expansion of the original statistic D. Next, we obtain transformed statistic $\tilde{D}$ by performing Bartlett-type transformation to D on the basis of the asymptotic expansion.

We will introduce some studies on asymptotic expansion for probability of a multinomial model. Regarding the goodness-of-fit test for a multinomial distribution, Yarnold [17] 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 (Siotani and Fujikoshi [12], Read [10], Menéndez et al. [8]). Edgeworth approximations of the distributions of some kinds of multinomial goodness-of-fit statistics under alternative hypotheses have also been investigated (Taneichi et al. [13, 14], Sekiya and Taneichi [11]). Taneichi and Sekiya [15] discussed approximations for the distribution of ϕ-divergence statistics for the test of independence in $r \times s$ contingency tables. Taneichi and Sekiya [16] also discussed approximations of the distributions of test statistics for homogeneity of a product multinomial model.

In this paper, we investigate asymptotic approximation of the distribution of the statistic D given by (1.2) for testing the null hypothesis H_0 given by (1.3). In Section 2, we consider expression of asymptotic expansion for the distribution of D under H_0. 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 a Bartlett-type transformation for improving small-sample accuracy of the χ^2 approximation of the distribution of D under H_0. In Section 4, the performance of the Bartlett-type transformed statistic and that of the original statistic are investigated 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 D under H_0

First, we consider a local Edgeworth approximation for the probability of Y_α, $(\alpha = 1, \ldots, N)$ under null hypothesis H_0 given by (1.3). Let

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

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

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

where

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

If we consider only for a limiting distribution of D, we can discuss under the assumption given by (1.4). In this section, since we consider asymptotic expansion of the distribution of D, we need an assumption that states the way of converging n_α/n to μ_α more strictly than the assumption given by (1.4). Therefore, we consider the following Assumption 1 instead of the assumption given by (1.4).

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

Assumption 1 and the assumption given by (1.4) state condition that n_α/n does not converge to 0 for every α, $(\alpha = 1, \ldots, N)$. However, for real data analysis, n_α, $(\alpha = 1, \ldots, N)$ and n are finite. So, for real data analysis, Assumption 1 and the assumption given by (1.4) imply the condition that excludes the case $n_\alpha = 0$ for some subgroups α, $(\alpha = 1, \ldots, N)$. Therefore, the range of applications does not change even if we change the assumption given by (1.4) to Assumption 1.

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

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

$$\Pr\{\boldsymbol{W} = \boldsymbol{w} | H_0\} = \left(\prod_{\alpha=1}^{N} \frac{1}{\sqrt{n_\alpha}} \right) h(\boldsymbol{w}) \left\{ 1 + \frac{1}{\sqrt{n}} g_1(\boldsymbol{w}) + \frac{1}{n} g_2(\boldsymbol{w}) + \frac{1}{n\sqrt{n}} g_3(\boldsymbol{w}) + O(n^{-2}) \right\}, \tag{2.2}$$

where

$$h(\boldsymbol{w}) = (2\pi)^{-N/2} |\Omega|^{-1/2} \exp\left(-\frac{1}{2} \boldsymbol{w}' \Omega^{-1} \boldsymbol{w} \right), \tag{2.3}$$

$$\begin{aligned}
g_1(\boldsymbol{w}) &= -\frac{1}{2} \sum_{\alpha=1}^{N} \frac{1}{\sqrt{\mu_\alpha}} \frac{(1 - 2\pi_\alpha)}{\pi_\alpha(1 - \pi_\alpha)} w_\alpha + \frac{1}{6} \sum_{\alpha=1}^{N} \frac{1}{\sqrt{\mu_\alpha}} \frac{(1 - 2\pi_\alpha)}{\pi_\alpha^2 (1 - \pi_\alpha)^2} w_\alpha^3, \\
g_2(\boldsymbol{w}) &= \frac{1}{2} \{g_1(\boldsymbol{w})\}^2 - \frac{1}{12} \sum_{\alpha=1}^{N} \frac{1}{\mu_\alpha} \frac{(1 - \pi_\alpha + \pi_\alpha^2)}{\pi_\alpha(1 - \pi_\alpha)} + \frac{1}{4} \sum_{\alpha=1}^{N} \frac{1}{\mu_\alpha} \frac{(1 - 2\pi_\alpha + 2\pi_\alpha^2)}{\pi_\alpha^2 (1 - \pi_\alpha)^2} w_\alpha^2 \\
&\quad - \frac{1}{12} \sum_{\alpha=1}^{N} \frac{1}{\mu_\alpha} \frac{(1 - 3\pi_\alpha + 3\pi_\alpha^2)}{\pi_\alpha^3 (1 - \pi_\alpha)^3} w_\alpha^4, \\
g_3(\boldsymbol{w}) &= -\frac{1}{3} \{g_1(\boldsymbol{w})\}^3 + g_1(\boldsymbol{w}) g_2(\boldsymbol{w}) + \frac{1}{12} \sum_{\alpha=1}^{N} \frac{1}{\mu_\alpha \sqrt{\mu_\alpha}} \frac{(1 - 2\pi_\alpha)}{\pi_\alpha^2 (1 - \pi_\alpha)^2} w_\alpha \\
&\quad - \frac{1}{6} \sum_{\alpha=1}^{N} \frac{1}{\mu_\alpha \sqrt{\mu_\alpha}} \frac{(1 - 2\pi_\alpha)(1 - \pi_\alpha + \pi_\alpha^2)}{\pi_\alpha^3 (1 - \pi_\alpha)^3} w_\alpha^3 \\
&\quad + \frac{1}{20} \sum_{\alpha=1}^{N} \frac{1}{\mu_\alpha \sqrt{\mu_\alpha}} \frac{(1 - 2\pi_\alpha)(1 - 2\pi_\alpha + 2\pi_\alpha^2)}{\pi_\alpha^4 (1 - \pi_\alpha)^4} w_\alpha^5,
\end{aligned}$$

and

$$\Omega = \mathrm{diag}(\pi_1(1 - \pi_1), \ldots, \pi_N(1 - \pi_N)). \tag{2.4}$$

By considering the proof of Theorem 22.1 of Bhattacharya and Ranga Rao ([3], pp. 232-236), we can prove Lemma 1. Proof of Lemma 1 is shown in Appendix 1.

Next, we derive an approximation based on an asymptotic expansion for the distribution of D under H_0. We consider the following approximation for the distribution of D under H_0 corresponding to approximation (2.3) of Taneichi et al. [13] for the multinomial goodness-of-fit test.

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

where the $J_1^*(x)$ term is multivariate Edgeworth expansion assuming a continuous distribution and the $J_2^*(x)$ term, which corresponds to the K_2 term of Taneichi et al. [13] 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 $J_1^*(x)$ term, we obtain the following theorem.

Theorem 1: *Under Assumption* 1, *the* $J_1^*(x)$ *term is evaluated as*

$$J_1^*(x) = \Pr\{\chi^2_{N-p} \leq x\} + \frac{1}{n}\sum_{j=0}^{1} v_j \Pr\{\chi^2_{N-p+2j} \leq x\} + O(n^{-2}), \tag{2.5}$$

where

$$v_0 = \frac{1}{24}(-6A_1 + 4A_2 + 6A_3 - 9A_4 + 2B_1 + 3B_2),$$

$$v_1 = -v_0,$$

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

$$A_2 = \sum_{\alpha=1}^{N} \frac{(1 - 2\pi_\alpha)^2}{\mu_\alpha \pi_\alpha (1 - \pi_\alpha)},$$

$$A_3 = \sum_{\alpha=1}^{N} \mu_\alpha \pi_\alpha (1 - \pi_\alpha)(1 - 3\pi_\alpha + 3\pi_\alpha^2)\sigma_{\alpha\alpha}^2,$$

$$A_4 = \sum_{\alpha=1}^{N} \mu_\alpha \pi_\alpha (1 - \pi_\alpha)(1 - 2\pi_\alpha)^2 \sigma_{\alpha\alpha}^2,$$

$$B_1 = \sum_{\alpha=1}^{N}\sum_{\gamma=1}^{N} \mu_\alpha \pi_\alpha (1 - \pi_\alpha)(1 - 2\pi_\alpha)\mu_\gamma \pi_\gamma (1 - \pi_\gamma)(1 - 2\pi_\gamma)\sigma_{\alpha\gamma}^3,$$

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

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

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

$\kappa^{l,m}$ is (l,m) elements of the inverse matrix of $K = (\kappa_{l,m})$, and χ_f^2 denotes a chi-square random variable with degrees of freedom f.

Proof of Theorem 1 is shown in Appendix 2.

Next, we consider the $J_2^*(x)$ term. Let $U(x)$ be a set defined by

$$U(x) = \{\boldsymbol{w} = (w_1, \ldots, w_N)' : D(\boldsymbol{w}) \leq x\}. \tag{2.6}$$

Consider the sets $U_\gamma \subset R_{N-1}$, $(\gamma = 1, \ldots, N)$ and continuous functions $\eta_\gamma(\cdot)$ and $\theta_\gamma(\cdot)$, $(\gamma = 1, \ldots, N)$ on R_{N-1} into R_1 such that $U(x)$ defined by (2.6) is represented as

$$\begin{aligned} U(x) \quad = \quad & \{\boldsymbol{w} = (w_1, \ldots, w_N)' : \eta_\gamma(\tilde{\boldsymbol{w}}_\gamma) \leq w_\gamma \leq \theta_\gamma(\tilde{\boldsymbol{w}}_\gamma), \\ & \qquad \tilde{\boldsymbol{w}}_\gamma = (w_1, \ldots, w_{\gamma-1}, w_{\gamma+1}, \ldots, w_N)' \in U_\gamma\}. \end{aligned}$$

Then

$$\begin{aligned} J_2^*(x) = -\frac{1}{\sqrt{n}} \sum_{\gamma=1}^{N} n^{-(N-\gamma)/2} \sum_{w_{\gamma+1} \in L_{\gamma+1}} \cdots \sum_{w_N \in L_N} \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \chi_{U_\gamma}(\tilde{\boldsymbol{w}}_\gamma) \\ \times \left[S_1\left(\sqrt{n} w_\gamma + n\pi_\gamma\right) h(\boldsymbol{w})\right]_{\eta_\gamma(\tilde{\boldsymbol{w}}_\gamma)}^{\theta_\gamma(\tilde{\boldsymbol{w}}_\gamma)} dw_1 \cdots dw_{\gamma-1}, \end{aligned} \tag{2.7}$$

where

$$\begin{aligned} [F(\boldsymbol{w})]_{\eta_\gamma(\tilde{\boldsymbol{w}}_\gamma)}^{\theta_\gamma(\tilde{\boldsymbol{w}}_\gamma)} \quad = \quad & F(w_1, \ldots, w_{\gamma-1}, \theta_\gamma(\tilde{\boldsymbol{w}}_\gamma), w_{\gamma+1}, \ldots, w_N) \\ & - F(w_1, \ldots, w_{\gamma-1}, \eta_\gamma(\tilde{\boldsymbol{w}}_\gamma), w_{\gamma+1}, \ldots, w_N), \end{aligned}$$

$$L_\gamma = \left\{ w_\gamma : w_\gamma = \frac{y_\gamma - n_\gamma \pi_\gamma}{\sqrt{n_\gamma}}, \quad y_\gamma \text{ is a non-negative integer which satisfies } y_\gamma \leq n_\gamma \right\},$$

$$(\gamma = 1, \ldots, N), \tag{2.8}$$

$$S_1(u) = u - [u] - \frac{1}{2}, \tag{2.9}$$

$h(\cdot)$ being defined by (2.3), and $\chi_A(\cdot)$ is the indicate function of the set A. In order to evaluate the $J_2^*(x)$ term of the null distribution of the test statistics using the same method as that of Yarnold [17], it is necessary to show

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

where b is a constant. However, it is very difficult to show the above relation except when $h(\boldsymbol{w})$ is a constant. Therefore, unlike the null distribution of multinomial goodness-of-fit test statistics, we cannot obtain a simple form of approximation of $J_2^*(x)$ such as $\hat{K}_2$ given by (2.6) of Taneichi et al. [13]. By another method of Yarnold [17], $J_2^*(x)$ is evaluated as follows.

Theorem 2: *Under Assumption* 1, *the* $J_2^*(x)$ *term can be represented in the following form:*

$$J_2^*(x) = \left\{(2\pi)^N \prod_{\alpha=1}^{N} \pi_\alpha(1-\pi_\alpha)\right\}^{-1/2} (\Theta_1 + \Theta_2) - \Theta_3 + O(n^{-2}), \tag{2.10}$$

where

$$\Theta_1 = n^{-N/2} \sum_{w_1 \in L_1} \cdots \sum_{\substack{w_N \in L_N \\ \boldsymbol{w} \in U(x)}} \exp\left(-\frac{1}{2}\boldsymbol{w}'\Omega^{-1}\boldsymbol{w}\right),$$

$$\begin{aligned}
\Theta_2 \;=\; & \frac{1}{\sqrt{n}} \frac{1}{\pi_N(1-\pi_N)} \int \cdots \int_{U(x)} w_N S_1\left(\sqrt{n}w_N + n\pi_N\right) \exp\left(-\frac{1}{2}\boldsymbol{w}'\Omega^{-1}\boldsymbol{w}\right) d\boldsymbol{w} \\
& + \frac{1}{n} \frac{1}{\pi_{N-1}(1-\pi_{N-1})} \sum_{w_N \in L_N} \int \cdots \int_{G_N(w_N)} w_{N-1} S_1\left(\sqrt{n}w_{N-1} + n\pi_{N-1}\right) \\
& \qquad \times \exp\left(-\tfrac{1}{2}\boldsymbol{w}'\Omega^{-1}\boldsymbol{w}\right) dw_1 \cdots dw_{N-1} \\
& + \frac{1}{n\sqrt{n}} \frac{1}{\pi_{N-2}(1-\pi_{N-2})} \sum_{w_{N-1} \in L_{N-1}} \sum_{w_N \in L_N} \int \cdots \int_{G_{N-1,N}(w_{N-1},w_N)} \\
& \qquad \times S_1\left(\sqrt{n}w_{N-2} + n\pi_{N-2}\right) \exp\left(-\tfrac{1}{2}\boldsymbol{w}'\Omega^{-1}\boldsymbol{w}\right) dw_1 \cdots dw_{N-2},
\end{aligned}$$

$$\Theta_3 = \Pr\{\chi^2_{N-p} \le x\} + \frac{1}{n} \sum_{j=0}^{3} \zeta_j \Pr\{\chi^2_{N-p+2j} \le x\},$$

where

$$G_N(w_N) = \{(w_1, \ldots, w_{N-1})' : \boldsymbol{w} = (w_1, \ldots, w_{N-1}, w_N)' \in U(x)\},$$

$$G_{N-1,N}(w_{N-1}, w_N) = \{(w_1, \ldots, w_{N-2})' : \boldsymbol{w} = (w_1, \ldots, w_{N-2}, w_{N-1}, w_N)' \in U(x)\},$$

$$\zeta_0 = \frac{1}{24}(-\Gamma_3),$$

$$\zeta_1 = \frac{1}{24}(\Gamma_1 + \Gamma_2 + \Gamma_3),$$

$$\zeta_2 = \frac{1}{24}(-\Gamma_1 - 2\Gamma_2),$$

$$\zeta_3 = \frac{1}{24}\Gamma_2,$$

$$\Gamma_1 = -3(2A_1 + 2A_3 - 6A_4 - 4A_5 + 4A_6 + 2B_1 + 3B_2 + B_3 - 4B_4),$$

$$\Gamma_2 = 5A_2 + 9A_4 - 12A_6 - 2B_1 - 3B_2 - 3B_3 + 6B_4,$$

$$\Gamma_3 = -3(4A_3 - 9A_4 - 4A_5 + 4A_6 + 4B_1 + 4B_2 + B_3 - 4B_4),$$

$$A_5 = \sum_{\alpha=1}^{N}(1 - 3\pi_\alpha + 3\pi_\alpha^2)\sigma_{\alpha\alpha},$$

$$A_6 = \sum_{\alpha=1}^{N}(1 - 2\pi_\alpha)^2\sigma_{\alpha\alpha},$$

$$B_3 = \sum_{\alpha=1}^{N}\sum_{\gamma=1}^{N}(1 - 2\pi_\alpha)(1 - 2\pi_\gamma)\sigma_{\alpha\gamma},$$

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

with $d\boldsymbol{w} = dw_1 \cdots dw_N$, *and* $A_1,\ldots,A_4$, B_1, B_2, *and* $\sigma_{\alpha\gamma}$ *being given in Theorem 1.*

Proof of Theorem 2 is shown in Appendix 3. By Theorem 2, we find that the $J_2^*(x)$ term is very difficult to calculate in practice. Then, on the basis of numerical results showing that Edgeworth approximation assuming a continuous distribution performs better than χ^2 approximation for a multinomial goodness-of-fit test (Taneichi et al. [13, 14]) and a test of independence in $r \times s$ contingency tables (Taneichi and Sekiya [15]), we consider the use of $J_1^*(x)$ as an approximation for the distribution of D under H_0.

3. Transformed deviance statistic based on the $J_1^*(x)$ term

In this section, we construct a Bartlett-type transformation for improving the accuracy of χ^2 approximation of the distribution of D under H_0 when the distribution of D is approximated as $J_1^*(x)$. The relation between coefficients of asymptotic expansion of a random variable and Bartlett adjustment of the random variable is shown as follows (e.g., Fijikoshi [5, 6]).

Theorem 3: *Suppose that a nonnegative random variable* T *has an asymptotic expansion such that*

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

The coefficients a_0 and a_1 do not depend on the parameter $n > 0$ and must satisfy the relation $a_1 = -a_0$. Then for a transformed random variable T_1 defined by

$$T_1 = \left(1 + \frac{2a_0}{fn}\right) T, \tag{3.1}$$

it holds that

$$\Pr\{T_1 \leq x\} = \Pr\{\chi_f^2 \leq x\} + O(n^{-2}). \tag{3.2}$$

T_1 is known as Bartlett adjustment of T. Lawley [7], Barndorff-Nielsen and Cox [1], and Barndorff-Nielsen and Hall [2] discussed Bartlett adjustment for the log likelihood ratio statistic. Applying evaluation (2.5) given by Theorem 1 to Bartlett adjustment (3.1), we obtain the following Bartlett-type adjustment D^*.

$$D^* = \left\{1 + \frac{2v_0}{n(N-p)}\right\} D. \tag{3.3}$$

In Barndorff-Nielsen and Cox [1], the theory of Bartlett adjustment is discussed for the case in which the error term in (3.2) is not $O(n^{-2})$ but $O(n^{-3/2})$. In Theorem 1, we evaluated the $J_1^*(x)$ term up to order $n^{-3/2}$. Therefore, we can apply the continuous part of asymptotic expansion for the distribution of D to Theorem 3, which ensures better accuracy of approximation than the theory of Barndorff-Nielsen and Cox [1].

Practically, we may use estimate $\hat{v}$ which is obtained by substituting maximum likelihood estimate $\hat{\boldsymbol{\beta}}$ for the true value $\boldsymbol{\beta}$ in v_0. Therefore, we propose the following Bartlett-type (transformed deviance) statistic $\tilde{D}$.

$$\tilde{D} = \left\{1 + \frac{2\hat{v}}{n(N-p)}\right\} D. \tag{3.4}$$

4. Performance of transformed deviance statistic

We compare the performance of the transformed deviance statistic $\tilde{D}$ given by (3.4) and that of the original deviance statistic D given by (1.2). We consider the logistic regression model given by (1.1) with $p = 2$ and $x_{\alpha 1} = 1$ and $x_{\alpha 2} = x_\alpha^*$, $(\alpha = 1, \ldots, N)$. This model is used as a dose-response model. Let the true values of parameters β_1 and β_2 be β_1^* and β_2^*, respectively. Then, the true values of π_α, $(\alpha = 1, \ldots, N)$ are

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

We give a design matrix

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

and execute the following procedure.

For each α, we generate n_α, $(\alpha = 1, \ldots, N)$ binomial random numbers which are distributed according to $B(1, \pi_\alpha^*)$, $(\alpha = 1, \ldots, N)$. From them, we calculate the number $Y_\alpha, (\alpha = 1, \ldots, N)$ of successes and the maximum likelihood estimates $\hat{\beta}_1$ and $\hat{\beta}_2$ for the parameters β_1 and β_2. Using the estimates, we calculate the values $\pi_\alpha(\hat{\boldsymbol{\beta}})$, $(\alpha = 1, \ldots, N)$, where $\hat{\boldsymbol{\beta}} = (\hat{\beta}_1, \hat{\beta}_2)'$, and observed values of the statistics D and $\tilde{D}$. This process is repeated J times.

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

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

We investigate the performance of the following four cases when $N = 8$.

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

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

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

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

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

$$\boldsymbol{X} = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 2.85 & 3.05 & 3.85 & 4.25 & 4.65 & 4.85 & 5.25 & 5.45 \end{pmatrix}'.$$

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

$$\boldsymbol{X} = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 0.15 & 0.25 & 0.35 & 0.4 & 0.45 & 0.5 & 0.55 & 0.6 \end{pmatrix}'.$$

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

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

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

(C) $n_1 = n_2 = n_C,\ n_3 = n_4 = n_C + 5,\ n_5 = n_6 = n_C + 10,\ n_7 = n_8 = n_C + 15$.

The cases and samples are selected appropriately in order to make many situations. Let $|I|$ be the absolute value of I. We calculate the value of $|I|$ 100 times and put them $I^*(i), i = 1, \ldots, 100$, where the number of repetitions is $J = 1.0 \times 10^4$. Let $\bar{I}^* = \sum_{i=1}^{100} I^*(i)/100$ and let I_T be the true value of $|I|$. For an approximate 95 percent confidence interval for I_T, we consider

$$\left[\bar{I}^* - t_{99}(0.025)s/\sqrt{100}, \bar{I}^* + t_{99}(0.025)s/\sqrt{100}\right],$$

where $s^2 = \sum_{i=1}^{100}(I^*(i) - \bar{I}^*)^2/99$ and $t_{99}(0.025)$ denotes the upper 2.5 percentage point of a t-distribution with degrees of freedom 99. Fig. 1 shows values of $\bar{I}^*$ and 95 percent confidence interval for I_T for sample design (A) where $n_A = 5, 10, 15, 20, 30$ and significance level $\varepsilon = 0.01, 0.05, 0.1$ for cases (I)–(IV). Figs. 2 and 3 show values of $\bar{I}^*$ and 95 percent confidence interval for I_T for sample designs (B) and (C) where n_B and $n_C = 5$, 10, 15, 20.

From Figs. 1–3, we find the following results. For all cases and sample designs, performance of transformed statistic $\tilde{D}$ is better than that of original statistic D. For almost all cases and sample designs, the value of $|I|$ for $\tilde{D}$ is less than one-third of that for statistic D. As a result of comparison, we can say that statistic D is improved by the transformed deviance statistic $\tilde{D}$. This result indicates that the Bartlett-type statistic works well.

Next, we consider the power of statistics D and $\tilde{D}$. We consider an alternative model:

$$\pi_\alpha^* = \frac{\exp(\beta_1^* + \beta_2^* x_\alpha^*)}{1 + \exp(\beta_1^* + \beta_2^* x_\alpha^*)} + \delta_\alpha, \qquad (\alpha = 1, \ldots, 8), \tag{4.1}$$

where $(\delta_1, \delta_2, \delta_3, \delta_4, \delta_5, \delta_6, \delta_7, \delta_8) = (-0.1,\ 0.1,\ -0.1,\ 0.1,\ -0.1,\ 0.1,\ -0.1,\ 0.1)$.

We calculate the simulated power against the alternative model (4.1) by using simulated exact critical values of statistic D and statistic $\tilde{D}$. We calculate simulated power 100 times and put them $P(i), i = 1, \ldots, 100$, where the number of repetitions is $J = 1.0 \times 10^4$. We consider the average simulated power $\bar{P} = \sum_{i=1}^{100} P(i)/100$. Let P_T be the true value of power. In the same way as that for I_T, we can derive the 95 percent confidence interval for P_T. Figs. 4–6 show the average simulated power $\bar{P}$ and 95 percent confidence interval for P_T when sample designs correspond to Figs. 1–3.

From Figs. 4–6, we find that the power of $\tilde{D}$ is not so different from the power of D. This result was expected since D^* given by (3.3) and deviance statistic D have the same exact power, theoretically.

As a matter of course, we can construct $\tilde{D}$ for a general logit model (1.1) when $p \geq 3$. We consider the general logit model given by (1.1) with $p = 3$ and $x_{\alpha 1} = 1$, $(\alpha = 1, \ldots, N)$. Using the same procedure and index as those in the case of $p = 2$ and $x_{\alpha 1} = 1$, $(\alpha = 1, \ldots, N)$, we investigate the performance of the following three cases when $N = 8$.

(V) True parameters are $\beta_1^* = 3, \beta_2^* = -8, \beta_3^* = 1$, and a design matrix is

$$\boldsymbol{X} = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 0.15 & 0.25 & 0.35 & 0.4 & 0.45 & 0.5 & 0.55 & 0.6 \\ 0.3 & 0.35 & 0.4 & 0.45 & 0.5 & 0.55 & 0.6 & 0.65 \end{pmatrix}'.$$

(VI) True parameters are $\beta_1^* = -4, \beta_2^* = 1, \beta_3^* = 2$, and a design matrix is

$$\boldsymbol{X} = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 2.7 & 3.0 & 3.3 & 3.5 & 4.3 & 4.9 & 5.0 & 5.2 \\ 0.2 & 0.25 & 0.3 & 0.35 & 0.4 & 0.45 & 0.5 & 0.55 \end{pmatrix}'.$$

(VII) True parameters are $\beta_1^* = 2, \beta_2^* = 3, \beta_3^* = -5$, and a design matrix is

$$\boldsymbol{X} = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 0.15 & 0.25 & 0.35 & 0.4 & 0.45 & 0.5 & 0.55 & 0.6 \\ 0.2 & 0.25 & 0.3 & 0.35 & 0.4 & 0.45 & 0.5 & 0.55 \end{pmatrix}'.$$

We also consider the same sample designs (A), (B) and (C) in the case of $p = 2$. Fig. 7 shows values of $\bar{I}^*$ and 95 percent confidence interval for I_T for sample design (A) where $n_A = 10, 15, 20, 30$ and significance level $\varepsilon = 0.01, 0.05, 0.1$ for cases (V)–(VII). Figs. 8 and 9 show values of $\bar{I}^*$ and 95 percent confidence interval for I_T for sample designs (B) and (C) where n_B and $n_C = 10, 15, 20$.

From Figs. 7–9, we find that D is also improved by the transformed statistic $\tilde{D}$ in the case of model (1.1) with $p = 3$. This result indicates that the Bartlett-type statistic also works well when the dimension of the model increases.

5. Real data application

By applying the transformed statistic to real data, we discuss the importance of the proposed transformed statistics. We use data based on an experiment by Farmer et al. [4]. In the experiment, female mice were fed dietary concentrations of one of 0.0, 0.3, 0.35, 0.45, 0.6, 0.75, 1.0 or 1.5 parts per 10^4 of a carcinogen, 2-acetylaminofluorene (2-AAF). Table 1 shows the incidences of bladder neoplasms in mice observed for 33 months. In Table 1, covariate variable $x_\alpha, (\alpha = 1, \dots, 8)$ is 2-AAF measured in parts per 10^4, $n_\alpha, (\alpha = 1, \dots, 8)$ is the number of mice exposed and $y_\alpha, (\alpha = 1, \dots, 8)$ is incidence of neoplasms. The logistic regression model which we apply is given by (1.1) with $p = 2$ and $x_{\alpha 1} = 1$ and $x_{\alpha 2} = x_\alpha, (\alpha = 1, \dots, 8)$, that is,

$$\text{logit } \pi_\alpha = \beta_1 + \beta_2 x_\alpha, \ (\alpha = 1, \dots, 8).$$

We consider testing the null hypothesis H_0 given by (1.3) using the deviance statistic at the significance level of 0.1. The maximum likelihood estimates of the parameters β_1 and β_2 are $\hat{\beta}_1 = -7.432$ and $\hat{\beta}_2$=7.875, respectively. By using $\hat{\pi}_\alpha = \pi_\alpha(\hat{\boldsymbol{\beta}})$, $(\alpha = 1, \dots, N)$, we calculate the observed value of D and the observed value of $\tilde{D}$. The observed value of D is 11.450 and that of $\tilde{D}$ is 3.211. The nominal critical value of significance level of 0.1 by using a chi-squared distribution is $\chi^2_6(0.1) = 10.645$. Then, if we use deviance statistic D, H_0 is rejected at the significance level of 0.1. However, if we use transformed statistic $\tilde{D}$, H_0 is accepted at the significance level of 0.1.

We consider the distribution of statistic D where D is constructed by random variable Y_α, $(\alpha = 1, \dots, N)$, provided that Y_α, $(\alpha = 1, \dots, N)$ is independently distributed according to the binomial distribution $B(n_\alpha, \hat{\pi}_\alpha), (\alpha = 1, \ldots, N)$. Let $D(0.1)$ be the upper 0.1 point of the distribution of D. Then, by using $D(0.1)$, we can execute an exact test at a significance level of 0.1. Therefore, as an accurate approximation of $D(0, 1)$, we consider a simulated approximation of $D(0.1)$ as follows.

For each α, by generating binomial random numbers $n_\alpha, (\alpha = 1, \ldots, N)$ which are distributed according to $B(1, \hat{\pi}_\alpha), (\alpha = 1, \ldots, N)$, we obtain $y^*_\alpha, (\alpha = 1, \ldots, N)$, which are observed values of $Y_\alpha, (\alpha = 1, \ldots, N)$. From $y^*_\alpha, (\alpha = 1, \ldots, N)$, we calculate the maximum likelihood estimate of $\boldsymbol{\beta}$ and observed value of D. By repeating this process $J = 10^6$ times, we obtain J observed values $D_{(j)}, (j = 1, \ldots, J)$. By sorting $D_{(j)}$, $(j = 1, \ldots, J)$ in large order, we adopt $0.1 \times J = 10^5$ th value as an approximation of $D(0.1)$ and put it $D_S(0.1)$.

In these data, we obtain $D_S(0.1) = 25.655$. Since $D_S(0.1) > D$, the result of the test by using the simulated critical value is accepted at the significance level of 0.1. That is, the test using the nominal critical value leads to a conclusion opposite to that obtained by the test using the simulated critical value. This result occurs on account of poorness of approximation for the upper probability of the deviance statistic.

On the other hand, by calculating simulated approximation of $\tilde{D}(0.1)$ for these data in the same way as $D_S(0.1)$, we obtain $\tilde{D}_S(0.1) = 13.483$. Since $\tilde{D}_S(0.1) > \tilde{D}$, the result of the test by using the simulated critical value is also accepted at the significance level of 0.1. That is, the result of the test using the nominal critical value coincides with that of the test using the simulated critical value. The above results are summarized in Table 2. This is an example of an asymptotic test based on the proposed transformed statistic $\tilde{D}$ being more reliable than that based on deviance statistic D.

6. Concluding remarks

We have shown the derivation of an expression of asymptotic expansion for the distribution of deviance statistic D in a logistic regression model. Using the continuous term of the expression of approximation for the distribution of deviance statistic D under a null hypothesis, we propose a transformation of D that improves the speed of convergence to a chi-square limiting distribution. Numerical comparison shows that the transformed deviance statistic $\tilde{D}$ is effective for improving the speed of convergence. This improvement increases the reliability of the results of the asymptotic test.

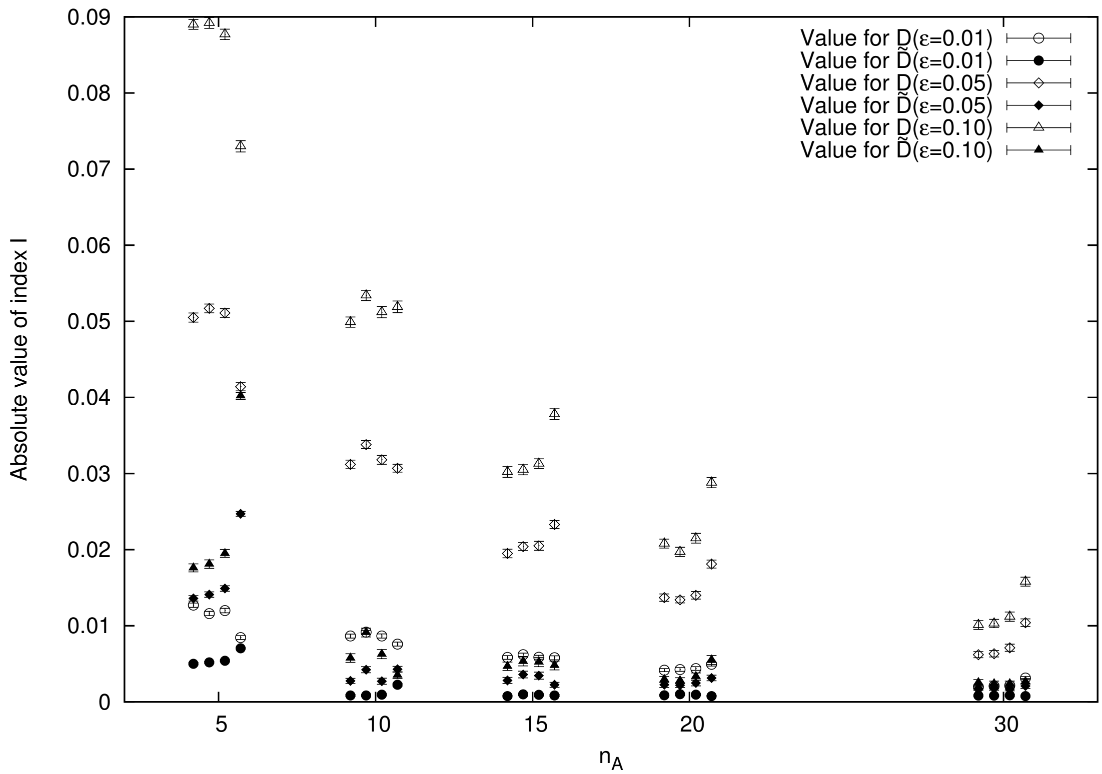

Fig. 1: $\bar{I}^*$ and 95 percent confidence interval for I_T for sample design (A), where $n_A = 5, 10, 15, 20, 30$: $\circ$, $\Diamond$ and $\triangle$ are the values for D when $\varepsilon = 0.01$, 0.05 and 0.1, respectively, and $\bullet$,$\blacklozenge$ and $\blacktriangle$ are the values for $\tilde{D}$ when $\varepsilon = 0.01$, 0.05 and 0.1, respectively: 1st column is for case (I), 2nd column is for case (II), 3rd column is for case (III), and 4th column is for case (IV).

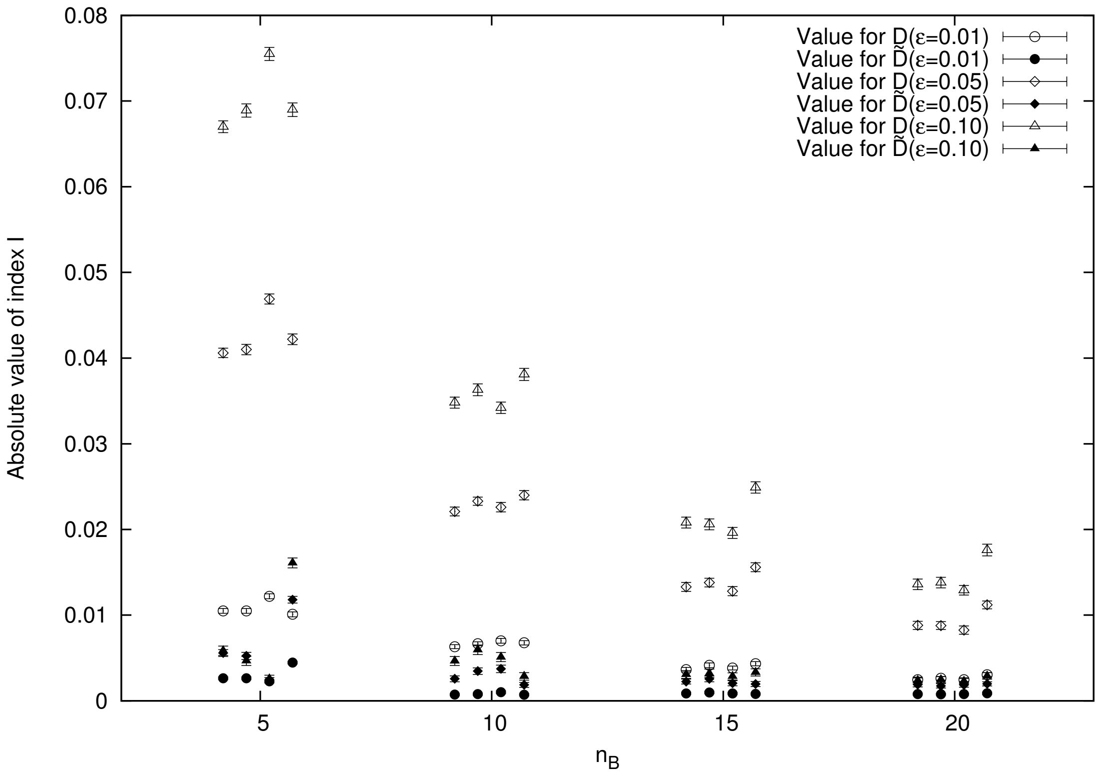

Fig. 2: $\bar{I}^*$ and 95 percent confidence interval for I_T for sample design (B), where $n_B = 5, 10, 15, 20$: $\circ$, $\Diamond$ and $\triangle$ are the values for D when $\varepsilon = 0.01$, 0.05 and 0.1, respectively, and $\bullet$,$\blacklozenge$ and $\blacktriangle$ are the values for $\tilde{D}$ when $\varepsilon = 0.01$, 0.05 and 0.1, respectively: 1st column is for case (I), 2nd column is for case (II), 3rd column is for case (III), and 4th column is for case (IV).

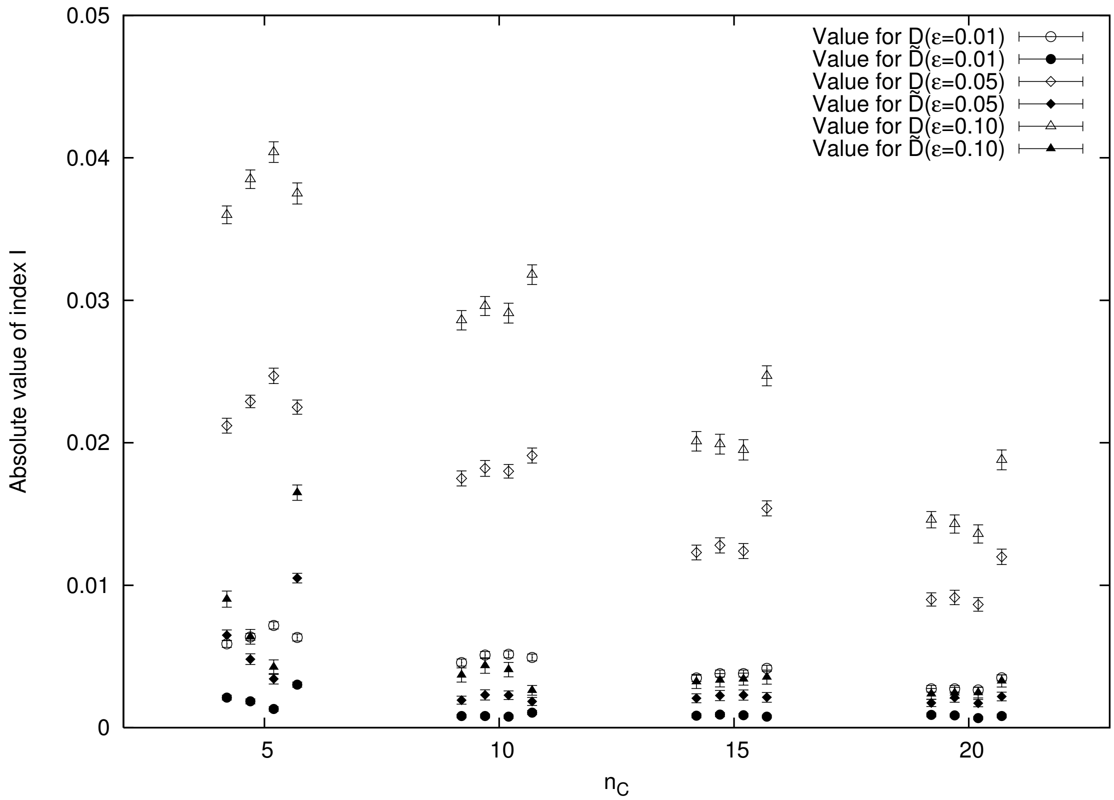

Fig. 3: $\bar{I}^*$ and 95 percent confidence interval for I_T for sample design (C), where $n_C = 5, 10, 15, 20$: $\circ$, $\Diamond$ and $\triangle$ are the values for D when $\varepsilon = 0.01$, 0.05 and 0.1, respectively, and $\bullet$,⧫ and ▲ are the values for $\tilde{D}$ when $\varepsilon = 0.01$, 0.05 and 0.1, respectively: 1st column is for case (I), 2nd column is for case (II), 3rd column is for case (III), and 4th column is for case (IV).

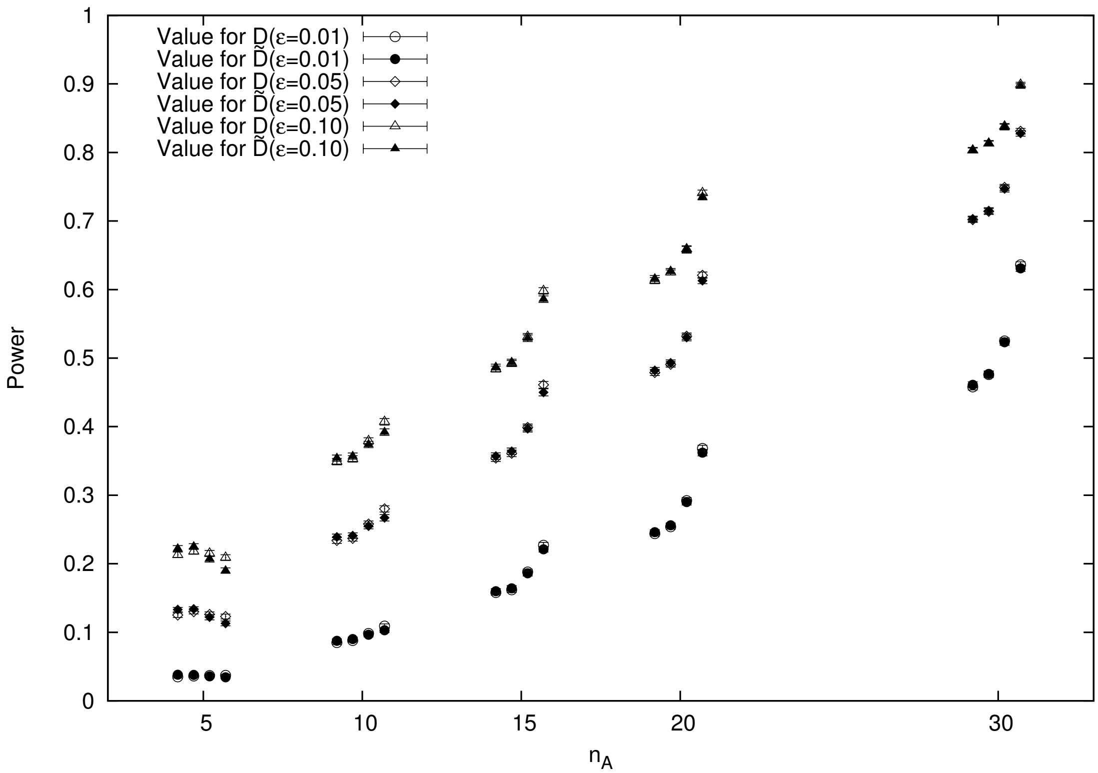

Fig. 4: $\bar{P}$ against alternative model (4.1) and 95 percent confidence interval for P_T for sample design (A), where $n_A = 5, 10, 15, 20, 30$: $\circ$, $\diamondsuit$ and $\triangle$ are the values for D when $\varepsilon = 0.01$, 0.05 and 0.1, respectively, and $\bullet$,$\blacklozenge$ and $\blacktriangle$ are the values for $\tilde{D}$ when $\varepsilon = 0.01$, 0.05 and 0.1, respectively: 1st column is for case (I), 2nd column is for case (II), 3rd column is for case (III), and 4th column is for case (IV).

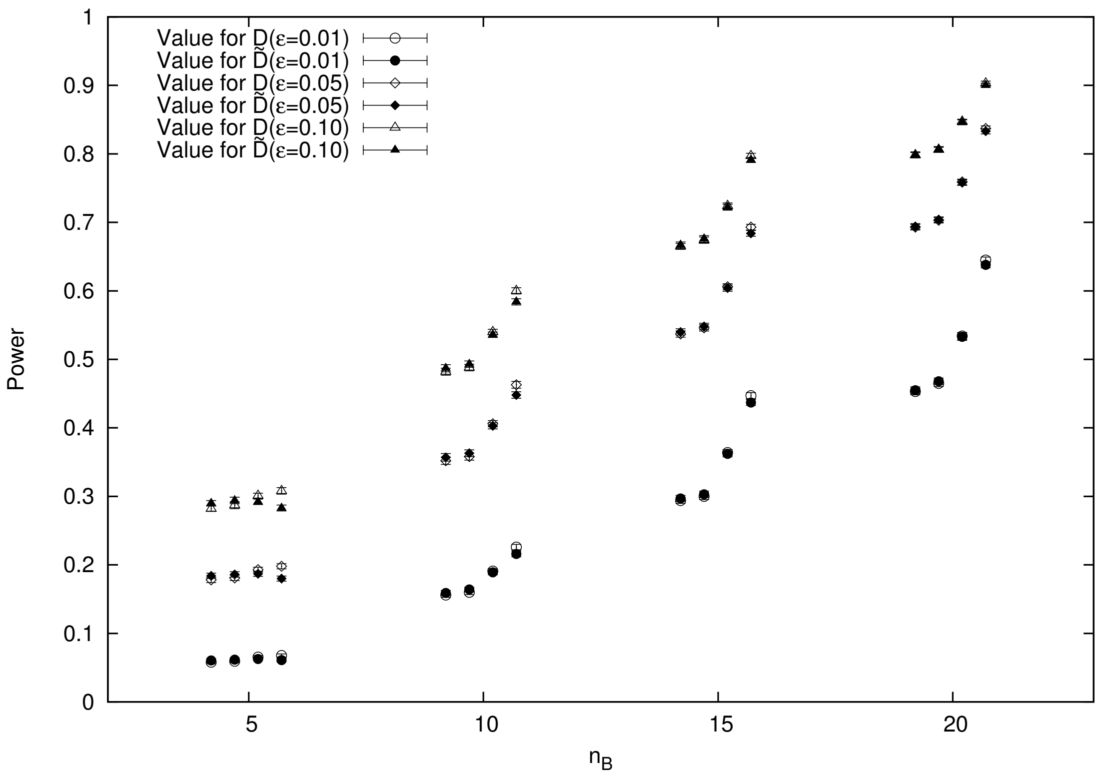

Fig. 5: $\bar{P}$ against alternative model (4.1) and 95 percent confidence interval for P_T for sample design (B), where $n_B = 5, 10, 15, 20$: $\circ$, $\Diamond$ and $\triangle$ are the values for D when $\varepsilon = 0.01$, 0.05 and 0.1, respectively, and $\bullet$,⧫ and ▲ are the values for $\tilde{D}$ when $\varepsilon = 0.01$, 0.05 and 0.1, respectively: 1st column is for case (I), 2nd column is for case (II), 3rd column is for case (III), and 4th column is for case (IV).

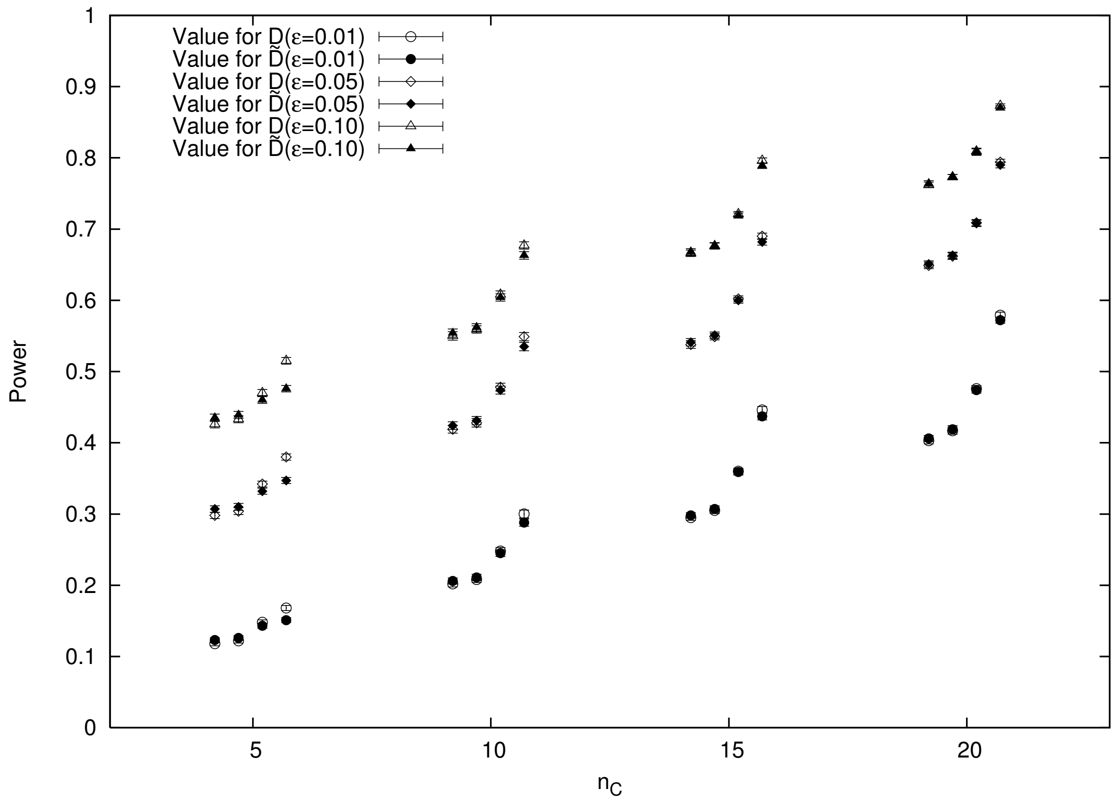

Fig. 6: $\bar{P}$ against alternative model (4.1) and 95 percent confidence interval for P_T for sample design (C), where $n_C = 5, 10, 15, 20$: $\circ$, $\Diamond$ and $\triangle$ are the values for D when $\varepsilon = 0.01$, 0.05 and 0.1, respectively, and $\bullet$,$\blacklozenge$ and $\blacktriangle$ are the values for $\tilde{D}$ when $\varepsilon = 0.01$, 0.05 and 0.1, respectively: 1st column is for case (I), 2nd column is for case (II), 3rd column is for case (III), and 4th column is for case (IV).

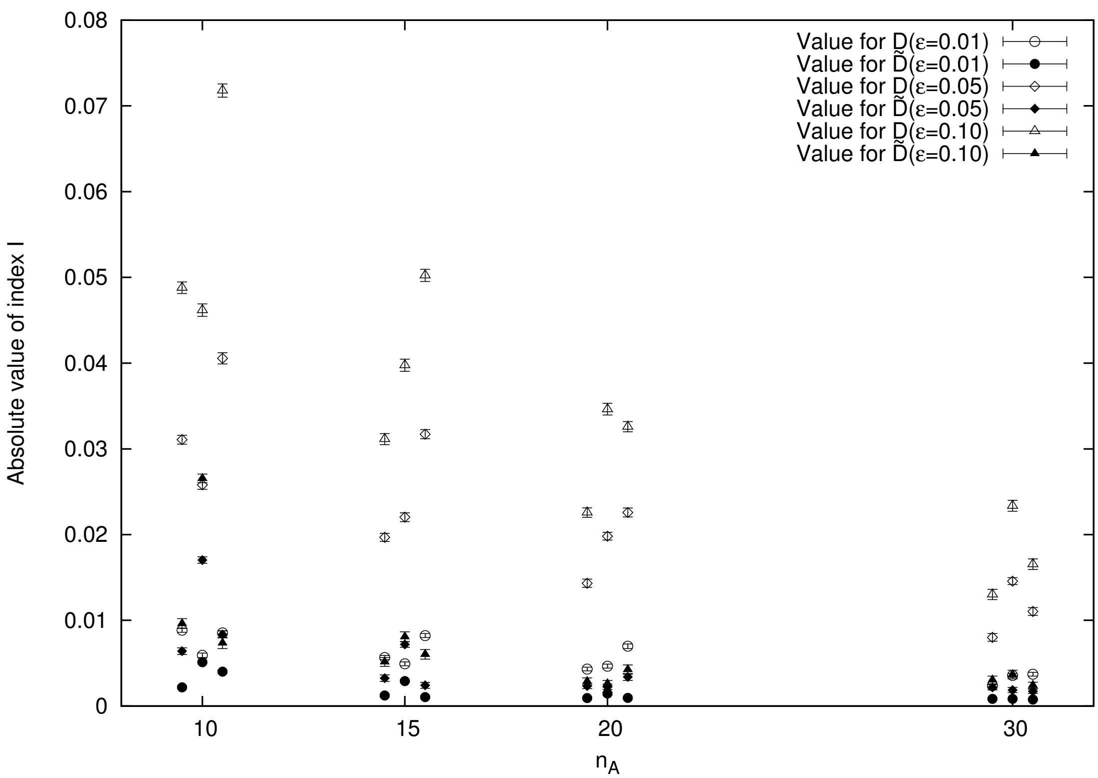

Fig. 7: $\bar{I}^*$ and 95 percent confidence interval for I_T for sample design (A), where $n_A = 10, 15, 20, 30$: $\circ$, $\Diamond$ and $\triangle$ are the values for D when $\varepsilon = 0.01$, 0.05 and 0.1, respectively, and $\bullet$,$\blacklozenge$ and $\blacktriangle$ are the values for $\tilde{D}$ when $\varepsilon = 0.01$, 0.05 and 0.1, respectively: 1st column is for case (V), 2nd column is for case (VI), and 3rd column is for case (VII).

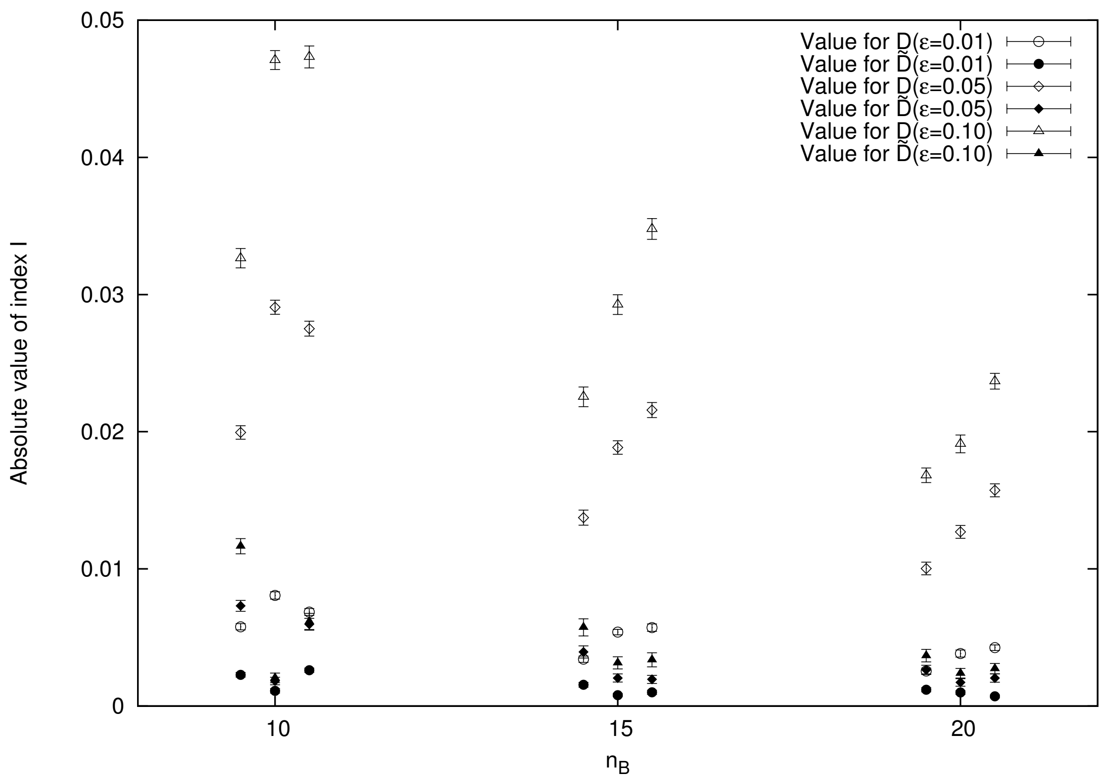

Fig. 8: $\bar{I}^*$ and 95 percent confidence interval for I_T for sample design (B), where $n_B = 10, 15, 20$: $\circ$, $\Diamond$ and $\triangle$ are the values for D when $\varepsilon = 0.01$, 0.05 and 0.1, respectively, and $\bullet$,$\blacklozenge$ and $\blacktriangle$ are the values for $\tilde{D}$ when $\varepsilon = 0.01$, 0.05 and 0.1, respectively: 1st column is for case (V), 2nd column is for case (VI), and 3rd column is for case (VII).

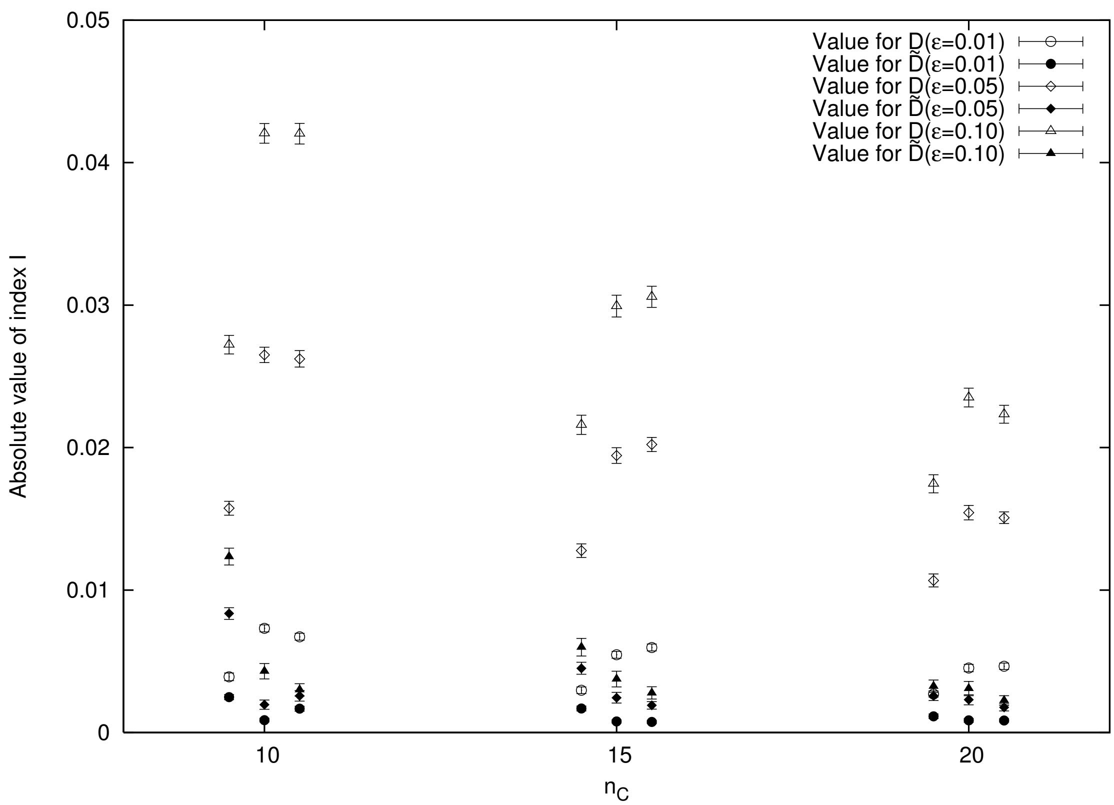

Fig. 9: $\bar{I}^*$ and 95 percent confidence interval for I_T for sample design (C), where $n_C = 10, 15, 20$: $\circ$, $\Diamond$ and $\triangle$ are the values for D when $\varepsilon = 0.01$, 0.05 and 0.1, respectively, and $\bullet$,$\blacklozenge$ and $\blacktriangle$ are the values for $\tilde{D}$ when $\varepsilon = 0.01$, 0.05 and 0.1, respectively: 1st column is for case (V), 2nd column is for case (VI), and 3rd column is for case (VII).

Table 1: Observed numbers of 2-AAF-exposed mice with bladder neoplasms.

α	Dose (parts per 10^4 2-AAF)(x_α)	Mice Exposed (n_α)	Incidence (y_α)
1	0.0	101	1
2	0.3	443	5
3	0.35	200	0
4	0.45	103	2
5	0.6	66	2
6	0.75	75	12
7	1.0	31	21
8	1.5	11	11

Table 2: Results of tests at the significance level of 0.1 based on simulated and nominal critical values for statistics D and $\tilde{D}$.

	D	$\tilde{D}$
observed value of test statistic	11.450	3.211
nominal critical value	10.645	10.645
result of test	reject	accept
simulated critical value	25.655	13.483
result of test	accept	accept

Appendix 1. (Proof of Lemma 1.)

Let $c(\boldsymbol{t})$ denote the characteristic function of $\boldsymbol{Y} = (Y_1, \ldots, Y_N)'$, where $\boldsymbol{t} = (t_1, \ldots, t_N)'$. Then

$$\begin{aligned} c(\boldsymbol{t}) &= \sum_{\boldsymbol{y} \in M} \exp(i\boldsymbol{t}'\boldsymbol{y}) \Pr\{\boldsymbol{Y} = \boldsymbol{y} \,|H_0\} \\ &= \prod_{\alpha=1}^{N} (\pi_\alpha e^{it_\alpha} + 1 - \pi_\alpha)^{n_\alpha}. \end{aligned}$$

For each $\boldsymbol{w} \in L$, we have

$$\begin{aligned} \Pr\{\boldsymbol{W} = \boldsymbol{w} \,|H_0\} &= \Pr\{\boldsymbol{Y} = \boldsymbol{y} \,|H_0\} \\ &= (2\pi)^{-N} \int_{-\pi}^{\pi} \cdots \int_{-\pi}^{\pi} c(\boldsymbol{t}) \exp(-i\boldsymbol{t}'\boldsymbol{y}) d\boldsymbol{t} \\ &= (2\pi)^{-N} \left(\prod_{\alpha=1}^{N} \frac{1}{\sqrt{n_\alpha}} \right) Q, \end{aligned}$$

where

$$Q = \int_{-\sqrt{n_1}\pi}^{\sqrt{n_1}\pi} \cdots \int_{-\sqrt{n_N}\pi}^{\sqrt{n_N}\pi} q(\boldsymbol{t}) \exp(-i\boldsymbol{t}'\boldsymbol{w}) d\boldsymbol{t}, \qquad (A1.1)$$

$$q(\boldsymbol{t}) = c(\boldsymbol{t}^*) \exp\left(-i \sum_{\alpha=1}^{N} \sqrt{n_\alpha} \pi_\alpha t_\alpha \right),$$

and

$$\boldsymbol{t}^* = \left(\frac{t_1}{\sqrt{n_1}}, \cdots, \frac{t_N}{\sqrt{n_N}} \right)'.$$

We can expand $q(\boldsymbol{t})$ as

$$q(\boldsymbol{t}) = \left\{ \exp\left(-\frac{1}{2} \boldsymbol{t}'\Omega\boldsymbol{t} \right) \right\} \left\{ 1 + \frac{1}{\sqrt{n}} b_1(\boldsymbol{t}) + \frac{1}{n} b_2(\boldsymbol{t}) + \frac{1}{n\sqrt{n}} b_3(\boldsymbol{t}) + O(n^{-2}) \right\} \qquad (A1.2)$$

for large n and fixed $\boldsymbol{t}$, where

$$b_1(\boldsymbol{t}) = \frac{i^3}{6} \sum_{\alpha=1}^{N} \frac{1}{\sqrt{\mu_\alpha}} \pi_\alpha (1 - \pi_\alpha)(1 - 2\pi_\alpha) t_\alpha^3,$$

$$b_2(\boldsymbol{t}) = \frac{1}{2} \{b_1(\boldsymbol{t})\}^2 + \frac{i^4}{24} \sum_{\alpha=1}^{N} \frac{1}{\mu_\alpha} \pi_\alpha (1 - \pi_\alpha)(1 - 6\pi_\alpha + 6\pi_\alpha^2) t_\alpha^4,$$

and

$$b_3(\boldsymbol{t}) = -\frac{1}{3} \{b_1(\boldsymbol{t})\}^3 + b_1(\boldsymbol{t}) b_2(\boldsymbol{t}) + \frac{i^5}{120} \sum_{\alpha=1}^{N} \frac{1}{\mu_\alpha \sqrt{\mu_\alpha}} \pi_\alpha (1 - \pi_\alpha)(1 - 2\pi_\alpha)(1 - 12\pi_\alpha + 12\pi_\alpha^2) t_\alpha^5.$$

From (A1.1) and (A1.2), we obtain

$$Q = Q_1 + Q_2 - Q_3,$$

where

$$\begin{aligned} Q_1 &= \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \{\exp(-i\boldsymbol{t}'\boldsymbol{w})\} \left\{\exp\left(-\frac{1}{2}\boldsymbol{t}'\Omega\boldsymbol{t}\right)\right\} \\ &\qquad \times \left\{1 + \frac{1}{\sqrt{n}}b_1(\boldsymbol{t}) + \frac{1}{n}b_2(\boldsymbol{t}) + \frac{1}{n\sqrt{n}}b_3(\boldsymbol{t})\right\} d\boldsymbol{t}, \end{aligned}$$

$$Q_2 = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \{\exp(-i\boldsymbol{t}'\boldsymbol{w})\} \left\{\exp\left(-\frac{1}{2}\boldsymbol{t}'\Omega\boldsymbol{t}\right)\right\} O(n^{-2}) d\boldsymbol{t},$$

$$\begin{aligned} Q_3 &= \int \cdots \int_{S^c} \{\exp(-i\boldsymbol{t}'\boldsymbol{w})\} \left\{\exp\left(-\frac{1}{2}\boldsymbol{t}'\Omega\boldsymbol{t}\right)\right\} \\ &\qquad \times \left\{1 + \frac{1}{\sqrt{n}}b_1(\boldsymbol{t}) + \frac{1}{n}b_2(\boldsymbol{t}) + \frac{1}{n\sqrt{n}}b_3(\boldsymbol{t}) + O(n^{-2})\right\} d\boldsymbol{t}, \end{aligned}$$

and

$$S = [-\sqrt{n_1}\pi, \sqrt{n_1}\pi] \times \cdots \times [-\sqrt{n_N}\pi, \sqrt{n_N}\pi].$$

Since $Q_2 = O(n^{-2})$ and $Q_3 = o(n^{-2})$, we obtain $Q = Q_1 + O(n^{-2})$. Therefore, we have

$$\begin{aligned} \Pr\{\boldsymbol{W} = \boldsymbol{w} | H_0\} &= (2\pi)^{-N} \left(\prod_{\alpha=1}^{N} \frac{1}{\sqrt{n_\alpha}}\right) \left[\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \{\exp(-i\boldsymbol{t}'\boldsymbol{w})\} \left\{\exp\left(-\frac{1}{2}\boldsymbol{t}'\Omega\boldsymbol{t}\right)\right\} \right. \\ &\qquad \left. \times \left\{1 + \frac{1}{\sqrt{n}}b_1(\boldsymbol{t}) + \frac{1}{n}b_2(\boldsymbol{t}) + \frac{1}{n\sqrt{n}}b_3(\boldsymbol{t})\right\} d\boldsymbol{t} + O(n^{-2})\right]. \end{aligned}$$

By carrying out this integration, we have (2.2). We have completed the proof of Lemma 1.

Appendix 2. (Proof of Theorem 1.)

By transformation (2.1), statistic D can be rewritten as

$$\begin{aligned} D(\boldsymbol{W}) &= 2\sum_{\alpha=1}^{N} n_\alpha \Bigg\{ \left(\pi_\alpha + W_\alpha(\sqrt{n_\alpha})^{-1}\right) \log\left(\frac{\pi_\alpha + W_\alpha(\sqrt{n_\alpha})^{-1}}{\hat{\pi}_\alpha(\boldsymbol{W})}\right) \\ &\qquad + \left(1 - \pi_\alpha - W_\alpha(\sqrt{n_\alpha})^{-1}\right) \log\left(\frac{1 - \pi_\alpha - W_\alpha(\sqrt{n_\alpha})^{-1}}{1 - \hat{\pi}_\alpha(\boldsymbol{W})}\right) \Bigg\}. \end{aligned}$$

If we regard

$$h(\boldsymbol{w}) \left\{1 + \frac{1}{\sqrt{n}}g_1(\boldsymbol{w}) + \frac{1}{n}g_2(\boldsymbol{w}) + \frac{1}{n\sqrt{n}}g_3(\boldsymbol{w})\right\}$$

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

$$J_1^*(x) = \int \cdots \int_{U(x)} h(\boldsymbol{w}) \left\{1 + \frac{1}{\sqrt{n}}g_1(\boldsymbol{w}) + \frac{1}{n}g_2(\boldsymbol{w}) + \frac{1}{n\sqrt{n}}g_3(\boldsymbol{w})\right\} d\boldsymbol{w}$$

as the distribution function of $D(\boldsymbol{W})$, where $U(x)$ is defined by (2.6). So, the characteristic function of $D(\boldsymbol{W})$ is calculated as

$$\psi(u)=\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}\left[\exp\{iuD(\boldsymbol{w})\}\right]h(\boldsymbol{w})\left\{1+\frac{1}{\sqrt{n}}g_1(\boldsymbol{w})+\frac{1}{n}g_2(\boldsymbol{w})+\frac{1}{n\sqrt{n}}g_3(\boldsymbol{w})\right\}d\boldsymbol{w}.$$

We can expand $D(\boldsymbol{w})$ as

$$D(\boldsymbol{w})=\tau_0(\boldsymbol{w})+\frac{1}{\sqrt{n}}\tau_1(\boldsymbol{w})+\frac{1}{n}\tau_2(\boldsymbol{w})+\frac{1}{n\sqrt{n}}\tau_3(\boldsymbol{w})+O(n^{-2}), \qquad (A2.1)$$

where

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

$$\begin{aligned}
\tau_1(\boldsymbol{w}) &= \frac{1}{3}\sum_{\alpha=1}^{N}\mu_\alpha\pi_\alpha(1-\pi_\alpha)(1-2\pi_\alpha)\left\{\sum_{l=1}^{p}x_{\alpha l}C_{1(l)}(\boldsymbol{w})\right\}^3-\frac{1}{3}\sum_{\alpha=1}^{N}\frac{1}{\sqrt{\mu_\alpha}}\frac{(1-2\pi_\alpha)}{\pi_\alpha^2(1-\pi_\alpha)^2}w_\alpha^3,\\
\tau_2(\boldsymbol{w}) &= \sum_{\alpha=1}^{N}\mu_\alpha\pi_\alpha(1-\pi_\alpha)\left\{\sum_{l=1}^{p}x_{\alpha l}C_{2(l)}(\boldsymbol{w})\right\}^2+\frac{1}{6}\sum_{\alpha=1}^{N}\frac{1}{\mu_\alpha}\frac{(1-3\pi_\alpha+3\pi_\alpha^2)}{\pi_\alpha^3(1-\pi_\alpha)^3}w_\alpha^4\\
&\quad+\sum_{\alpha=1}^{N}\mu_\alpha\pi_\alpha(1-\pi_\alpha)(1-2\pi_\alpha)\left\{\sum_{l=1}^{p}x_{\alpha l}C_{1(l)}(\boldsymbol{w})\right\}^2\left\{\sum_{l=1}^{p}x_{\alpha l}C_{2(l)}(\boldsymbol{w})\right\}\\
&\quad+\frac{1}{12}\sum_{\alpha=1}^{N}\mu_\alpha\pi_\alpha(1-\pi_\alpha)(1-6\pi_\alpha+6\pi_\alpha^2)\left\{\sum_{l=1}^{p}x_{\alpha l}C_{1(l)}(\boldsymbol{w})\right\}^4,\\
\tau_3(\boldsymbol{w}) &= 2\sum_{\alpha=1}^{N}\mu_\alpha\pi_\alpha(1-\pi_\alpha)\left\{\sum_{l=1}^{p}x_{\alpha l}C_{2(l)}(\boldsymbol{w})\right\}\left\{\sum_{l=1}^{p}x_{\alpha l}C_{3(l)}(\boldsymbol{w})\right\}\\
&\quad+\sum_{\alpha=1}^{N}\mu_\alpha\pi_\alpha(1-\pi_\alpha)(1-2\pi_\alpha)\left\{\sum_{l=1}^{p}x_{\alpha l}C_{1(l)}(\boldsymbol{w})\right\}^2\left\{\sum_{l=1}^{p}x_{\alpha l}C_{3(l)}(\boldsymbol{w})\right\}\\
&\quad+\sum_{\alpha=1}^{N}\mu_\alpha\pi_\alpha(1-\pi_\alpha)(1-2\pi_\alpha)\left\{\sum_{l=1}^{p}x_{\alpha l}C_{1(l)}(\boldsymbol{w})\right\}\left\{\sum_{l=1}^{p}x_{\alpha l}C_{2(l)}(\boldsymbol{w})\right\}^2\\
&\quad+\frac{1}{3}\sum_{\alpha=1}^{N}\mu_\alpha\pi_\alpha(1-\pi_\alpha)(1-6\pi_\alpha+6\pi_\alpha^2)\left\{\sum_{l=1}^{p}x_{\alpha l}C_{1(l)}(\boldsymbol{w})\right\}^3\left\{\sum_{l=1}^{p}x_{\alpha l}C_{2(l)}(\boldsymbol{w})\right\}\\
&\quad+\frac{1}{60}\sum_{\alpha=1}^{N}\mu_\alpha\pi_\alpha(1-\pi_\alpha)(1-2\pi_\alpha)(1-12\pi_\alpha+12\pi_\alpha^2)\left\{\sum_{l=1}^{p}x_{\alpha l}C_{1(l)}(\boldsymbol{w})\right\}^5\\
&\quad-\frac{1}{10}\sum_{\alpha=1}^{N}\frac{1}{\mu_\alpha\sqrt{\mu_\alpha}}\frac{(1-2\pi_\alpha)(1-2\pi_\alpha+2\pi_\alpha^2)}{\pi_\alpha^4(1-\pi_\alpha)^4}w_\alpha^5,
\end{aligned}$$

$$\Xi=\begin{pmatrix}\sqrt{\mu_1\mu_1}\sigma_{11} & \cdots & \sqrt{\mu_1\mu_N}\sigma_{1N}\\ \vdots & \ddots & \vdots\\ \sqrt{\mu_N\mu_1}\sigma_{N1} & \cdots & \sqrt{\mu_N\mu_N}\sigma_{NN}\end{pmatrix}, \qquad (A2.2)$$

$$C_{1(l)}(\boldsymbol{w}) = \sum_{m=1}^{p} \kappa^{l,m} \varphi_m(\boldsymbol{w}),\ (l = 1, \ldots, p),$$

$$C_{2(l)}(\boldsymbol{w}) = -\frac{1}{2} \sum_{m_1=1}^{p} \cdots \sum_{m_5=1}^{p} \kappa^{l,m_3} \kappa^{m_1,m_4} \kappa^{m_2,m_5} \kappa_{m_3,m_4,m_5} \varphi_{m_1}(\boldsymbol{w}) \varphi_{m_2}(\boldsymbol{w}),\ (l = 1, \ldots, p),$$

$$\begin{aligned} C_{3(l)}(\boldsymbol{w}) &= \frac{1}{2} \sum_{m_1=1}^{p} \cdots \sum_{m_9=1}^{p} \kappa^{l,m_4} \kappa^{m_1,m_5} \kappa^{m_2,m_6} \kappa^{m_3,m_7} \kappa^{m_8,m_9} \kappa_{m_4,m_5,m_8} \kappa_{m_6,m_7,m_9} \\ &\qquad \times \varphi_{m_1}(\boldsymbol{w}) \varphi_{m_2}(\boldsymbol{w}) \varphi_{m_3}(\boldsymbol{w}) \\ &\quad -\frac{1}{6} \sum_{m_1=1}^{p} \cdots \sum_{m_7=1}^{p} \kappa^{l,m_4} \kappa^{m_1,m_5} \kappa^{m_2,m_6} \kappa^{m_3,m_7} \kappa_{m_4,m_5,m_6,m_7} \\ &\qquad \times \varphi_{m_1}(\boldsymbol{w}) \varphi_{m_2}(\boldsymbol{w}) \varphi_{m_3}(\boldsymbol{w}), \qquad (l = 1, \ldots, p), \end{aligned}$$

$$\kappa_{m_1,m_2,m_3} = \sum_{\lambda=1}^{N} \mu_\lambda \pi_\lambda (1 - \pi_\lambda)(1 - 2\pi_\lambda) x_{\lambda m_1} x_{\lambda m_2} x_{\lambda m_3},\ (m_1, m_2, m_3 = 1, \ldots, p),$$

$$\begin{aligned} \kappa_{m_1,m_2,m_3,m_4} &= \sum_{\lambda=1}^{N} \mu_\lambda \pi_\lambda (1 - \pi_\lambda)(1 - 6\pi_\lambda + 6\pi_\lambda^2) x_{\lambda m_1} x_{\lambda m_2} x_{\lambda m_3} x_{\lambda m_4}, \\ &\qquad (m_1, m_2, m_3, m_4 = 1, \ldots, p), \end{aligned}$$

$$\varphi_m(\boldsymbol{w}) = \sum_{\lambda=1}^{N} \sqrt{\mu_\lambda} x_{\lambda m} w_\lambda,\ (m = 1, \ldots, p),$$

Ω is defined by (2.4), and $\sigma_{\alpha\beta}$ and $\kappa^{l,m}$ are defined in Theorem 1. Then from (A2.1), we obtain

$$\begin{aligned} &[\exp\{iuD(\boldsymbol{w})\}]\, h(\boldsymbol{w}) \left\{1 + \frac{1}{\sqrt{n}} g_1(\boldsymbol{w}) + \frac{1}{n} g_2(\boldsymbol{w}) + \frac{1}{n\sqrt{n}} g_3(\boldsymbol{w})\right\} \\ = \ &(2\pi)^{-N/2} |\Omega|^{-1/2} \left[\exp\left\{-\frac{1}{2} \boldsymbol{w}' \left((1 - 2iu)\Omega^{-1} + 2iu\Xi\right) \boldsymbol{w}\right\}\right] \left[G(\boldsymbol{w}) + O(n^{-2})\right], \quad (A2.3) \end{aligned}$$

where

$$\begin{aligned} G(\boldsymbol{w}) &= 1 + \frac{1}{\sqrt{n}} \{g_1(\boldsymbol{w}) + (iu)\tau_1(\boldsymbol{w})\} \\ &\quad + \frac{1}{n} \left[g_2(\boldsymbol{w}) + (iu)\tau_1(\boldsymbol{w}) g_1(\boldsymbol{w}) + (iu)\tau_2(\boldsymbol{w}) + \frac{1}{2}(iu)^2 \{\tau_1(\boldsymbol{w})\}^2\right] \\ &\quad + \frac{1}{n\sqrt{n}} \Big[g_3(\boldsymbol{w}) + (iu)\tau_1(\boldsymbol{w}) g_2(\boldsymbol{w}) + (iu)\tau_2(\boldsymbol{w}) g_1(\boldsymbol{w}) + \frac{1}{2}(iu)^2 \{\tau_1(\boldsymbol{w})\}^2 g_1(\boldsymbol{w}) \\ &\qquad + (iu)\tau_3(\boldsymbol{w}) + (iu)^2 \tau_1(\boldsymbol{w}) \tau_2(\boldsymbol{w}) + \frac{1}{6}(iu)^3 \{\tau_1(\boldsymbol{w})\}^3\Big]. \end{aligned}$$

Let

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

where Ω and Ξ are defined by (2.4) and (A2.2), respectively. Then

$$\Lambda^{-1} = (1-2iu)\Omega^{-1} + 2iu\Xi \qquad (A2.4)$$

and

$$|\Lambda| = (1-2iu)^{-(N-p)}|\Omega|. \qquad (A2.5)$$

Therefore, from (A2.3), (A2.4) and (A2.5), we obtain

$$\psi(u) = (1-2iu)^{-(N-p)/2}\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}(2\pi)^{-N/2}|\Lambda|^{-1/2}\left\{\exp\left(-\frac{1}{2}\boldsymbol{w}'\Lambda^{-1}\boldsymbol{w}\right)\right\}G(\boldsymbol{w})d\boldsymbol{w} + O(n^{-2}). \qquad (A2.6)$$

Since $\tau_j(\boldsymbol{w}), (j=1,2,3)$ is a homogeneous polynomial of degree $j+2$ with respect to variable $w_1,\ldots,w_N$, and degrees of all terms of polynomial $g_j(\boldsymbol{w}), (j=1,2,3)$ are odd if $j=1$ or 3 and even if $j=2$, degrees of all terms of polynomial $G(\boldsymbol{w})$ for order $n^{-1/2}$ and $n^{-3/2}$ are odd. Therefore, by carrying out the integration of (A2.6), the characteristic function $\psi(u)$ is expanded as

$$\psi(u) = (1-2iu)^{-(N-p)/2}\left[1+\frac{1}{n}\sum_{j=0}^{1}(1-2iu)^{-j}v_j + O(n^{-2})\right]. \qquad (A2.7)$$

Since $(1-2iu)^{-(N-p)/2}$ is the characteristic function of the χ^2_{N-p} distribution, by inverting (A2.7), we obtain (2.5). We have completed the proof of Theorem 1.

Appendix 3. (Proof of Theorem 2.)

The function $S_1(\sqrt{n}w_\gamma + n\pi_\gamma)$ defined by (2.9) is differentiable except when $w_\gamma \in L_\gamma$ defined by (2.8), and $h(\boldsymbol{w})$ is a differentiable function on R_N. Therefore, by (8.10) in the proof of Lemma 1 of Yarnold [17],

$$\left[S_1\left(\sqrt{n}w_\gamma + n\pi_\gamma\right)h(\boldsymbol{w})\right]_{\eta_\gamma(\tilde{\boldsymbol{w}}_\gamma)}^{\theta_\gamma(\tilde{\boldsymbol{w}}_\gamma)} = \int_{\eta_\gamma(\tilde{\boldsymbol{w}}_\gamma)}^{\theta_\gamma(\tilde{\boldsymbol{w}}_\gamma)} D_\gamma S_1\left(\sqrt{n}w_\gamma + n\pi_\gamma\right)h(\boldsymbol{w})dw_\gamma + \sum_{\substack{w_\gamma=\eta_\gamma(\tilde{\boldsymbol{w}}_\gamma)\\ w_\gamma\in L_\gamma}}^{\theta_\gamma(\tilde{\boldsymbol{w}}_\gamma)} \Delta_\gamma S_1\left(\sqrt{n}w_\gamma + n\pi_\gamma\right)h(\boldsymbol{w}), \qquad (A3.1)$$

where

$$\Delta_\gamma F(\boldsymbol{w}) = F(w_1,\ldots,w_{\gamma-1},w_\gamma+0,w_{\gamma+1},\ldots,w_N) - F(w_1,\ldots,w_{\gamma-1},w_\gamma-0,w_{\gamma+1},\ldots,w_N)$$

and $D_\gamma F(\boldsymbol{w}) = (\partial/\partial w_\gamma)F(\boldsymbol{w})$. By definitions of the functions $S_1(\cdot)$ and $h(\cdot)$, we obtain

$$\Delta_\gamma S_1\left(\sqrt{n}w_\gamma + n\pi_\gamma\right) h(\boldsymbol{w}) = -h(\boldsymbol{w}), \tag{A3.2}$$

$$D_\gamma S_1\left(\sqrt{n}w_\gamma + n\pi_\gamma\right) h(\boldsymbol{w}) = h(\boldsymbol{w})\left\{\sqrt{n} - S_1\left(\sqrt{n}w_\gamma + n\pi_\gamma\right)\frac{w_\gamma}{\pi_\gamma(1-\pi_\gamma)}\right\}, \tag{A3.3}$$

and

$$h(\boldsymbol{w}) = \left\{(2\pi)^N \prod_{\alpha=1}^{N} \pi_\alpha(1-\pi_\alpha)\right\}^{-1/2} \exp\left(-\frac{1}{2}\boldsymbol{w}'\Omega^{-1}\boldsymbol{w}\right). \tag{A3.4}$$

By substituting (A3.2), (A3.3) and (A3.4) for (A3.1), we obtain the following:

$$\begin{aligned}
&\left[S_1\left(\sqrt{n}w_\gamma + n\pi_\gamma\right) h(\boldsymbol{w})\right]_{\eta_\gamma(\tilde{\boldsymbol{w}}_\gamma)}^{\theta_\gamma(\tilde{\boldsymbol{w}}_\gamma)} = \left\{(2\pi)^N \prod_{\alpha=1}^{N} \pi_\alpha(1-\pi_\alpha)\right\}^{-1/2} \\
&\times\left\{\sqrt{n}\int_{\eta_\gamma(\tilde{\boldsymbol{w}}_\gamma)}^{\theta_\gamma(\tilde{\boldsymbol{w}}_\gamma)} \exp\left(-\frac{1}{2}\boldsymbol{w}'\Omega^{-1}\boldsymbol{w}\right) dw_\gamma - \sum_{\substack{w_\gamma=\eta_\gamma(\tilde{\boldsymbol{w}}_\gamma)\\ w_\gamma\in L_\gamma}}^{\theta_\gamma(\tilde{\boldsymbol{w}}_\gamma)} \exp\left(-\frac{1}{2}\boldsymbol{w}'\Omega^{-1}\boldsymbol{w}\right)\right. \\
&\left. -\frac{1}{\pi_\gamma(1-\pi_\gamma)}\int_{\eta_\gamma(\tilde{\boldsymbol{w}}_\gamma)}^{\theta_\gamma(\tilde{\boldsymbol{w}}_\gamma)} w_\gamma S_1\left(\sqrt{n}w_\gamma + n\pi_\gamma\right)\exp\left(-\frac{1}{2}\boldsymbol{w}'\Omega^{-1}\boldsymbol{w}\right) dw_\gamma\right\}.
\end{aligned} \tag{A3.5}$$

Then from (2.7) and (A3.5), we obtain the following:

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

where

$$\Theta_3^* = \int\cdots\int_{U(x)} h(\boldsymbol{w})d\boldsymbol{w}$$

and

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

If we regard $h(\boldsymbol{w})$ as the density function of $\boldsymbol{W}$, then we can regard Θ_3^* as the distribution function of $D(\boldsymbol{W})$. Then, by expanding the characteristic function of $D(\boldsymbol{W})$ and inverting it, we can approximate $\Theta_3^* = \Theta_3 + O(n^{-2})$. Furthermore, we can approximate $\Theta_2^* = \Theta_2 + O(n^{-2})$. Therefore, we obtain (2.10). We have completed the proof of Theorem 2.

References

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

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

[3] Bhattacharya, R. N. and Ranga Rao, R.: *Normal approximation and asymptotic expansion*, (1976), New York: Wiley.

[4] Farmer, J. H., Kodell, R. L. and Greenman, D. L.: Dose and time response models for the incidence of bladder and liver neoplasms in mice fed 2-acetylaminofluorene continuously. *Journal of Environmental Pathology and Toxicology*, **3**, (1979), 55–68.

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

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

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

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

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

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

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

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

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

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

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

[16] Taneichi, N. and Sekiya, Y.: Approximations of the distributions of test statistics for homogeneity of a product multinomial model, *Commun. Statist. - Theory and Meth.*, **37**(10), (2008), 1610–1631.

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