SUT Journal of Mathematics
Vol. 57, No. 2 (2021), 167–183

Improved approximate unbiased estimators of the measure of departure from partial symmetry for square contingency tables

Tomoyuki Nakagawa, Ryoma Namba, Kiyotaka Iki and Sadao Tomizawa

(Received May 11, 2021)

Abstract. For square contingency tables, the measure to represent the degree of departure from the partial symmetry model was proposed. It is necessary to estimate the measure because it is constructed of unknown parameters. Although many studies consider using the plug-in estimator to estimate the measure, the bias of the plug-in estimator is large when the sample size is not so large. In this study, we consider to estimate the measure when the sample size is not so large. This paper presents the improved approximate unbiased estimators of the measure which are obtained using the second-order term in Taylor series expansion. Some simulation studies show the performances of proposed estimators for finite sample cases.

AMS 2010 *Mathematics Subject Classification.* 62H17, 62F12.

Key words and phrases. Measure, Partial symmetry, Square contingency table, Symmetry, Unbiased estimator.

§1. Introduction

Consider an $r \times r$ square contingency table with the same row and column classifications. Let p_{ij} denote the probability that an observation will fall in the ith row and jth column of the table ($i = 1, \dots, r;\ j = 1, \dots, r$). Many statisticians are probably interested in symmetry rather than independence because row and column variables of such contingency tables are rarely independent. Bowker [3] proposed the symmetry (S) model defined by

$$p_{ij} = \psi_{ij}\ (i = 1, \dots, r;\ j = 1, \dots, r),$$

where $\psi_{ij} = \psi_{ji}$. For the analysis of data, the S model may fit the data poorly because it has a strong restriction. In such cases, many statisticians may be interested in applying some models which have weaker restrictions than the S model. Numerous studies have developed new models which have weaker restrictions than the S model (see [1, 2, 4, 5, 6]). Saigusa *et al.* [7] proposed the partial symmetry (PS) model defined by

$$p_{ij} = \psi_{ij} \ (i = 1, \dots, r; \ j = 1, \dots, r),$$

where $\psi_{st} = \psi_{ts}$ for at least one (s, t) with $s \neq t$. This model states that the probability that an observation will fall in row category s and column category t is equal to the probability that the observation falls in row category t and column category s for at least one (s, t), $s \neq t$. It is easy to know that the S model implies the PS model since the S model indicates that p_{ij} equals p_{ji} for all (i, j).

On the other hand, when the models fit the data poorly, we are also interested in measuring the degree of departure from the models. Tomizawa [10] and Tomizawa *et al.* [11] proposed the measure to represent the degree of departure from the S model for nominal data. Saigusa *et al.* [7] also proposed the measure to represent the degree of departure from the PS model for square contingency tables. Since the measures are constructed in the cell probabilities which are unknown parameters, many studies considered the plug-in estimator and derived its confidence interval for the measure. However, when the sample size is not so large, the bias of the plug-in estimator is large. To overcome this problem, Tomizawa *et al.* [12] proposed the improved approximate unbiased estimators of the measure for departure from the S model when the sample size is not so large. Moreover, using a similar technique, the improved approximate unbiased estimators of log-odds ratio and the measures for the marginal homogeneity model have been proposed in [8, 9].

The purpose of this paper is to propose the approximate unbiased estimators which are better than the plug-in estimator of the measure for departure from the PS model. The rest of this paper is organized as follows: In Section 2, we describe the measure to represent that degree of departure from the PS model. Section 3 proposes the improved approximate unbiased estimators of the measure for departure from the PS model for square contingency tables. In section 4, we compare the biases and the mean square errors of proposed estimators with that of [7] in simulation studies. Section 5 presents concluding remarks.

§2. Measures

Saigusa *et al.* [7] proposed the following measure $\Phi^{(\lambda)}$ to represent the degree of departure from the PS model for square contingency tables. Let $p_{ij} + p_{ji} \neq$

$0 \quad (i \neq j; i, j = 1, \ldots, r)$ and let

$$\delta = \sum_{s=1}^{r} \sum_{\substack{t=1 \\ t \neq s}}^{r} p_{st}, \quad p_{ij}^{*} = \frac{p_{ij}}{\delta}, \quad p_{ij}^{c} = \frac{p_{ij}}{p_{ij} + p_{ji}},$$

for $i = 1, \ldots, r;\ j = 1, \ldots, r;\ i \neq j$, then the measure $\Phi^{(\lambda)}$ is defined by

$$\Phi^{(\lambda)} = \prod_{i=1}^{r-1} \prod_{j=i+1}^{r} \left[1 - \frac{\lambda 2^{\lambda}}{2^{\lambda} - 1} H_{ij}^{(\lambda)}\right]^{(p_{ij}^{*} + p_{ji}^{*})} \quad \text{for } \lambda > -1,$$

where

$$H_{ij}^{(\lambda)} = \frac{1}{\lambda} \left[1 - \left(p_{ij}^{c}\right)^{\lambda+1} - \left(p_{ji}^{c}\right)^{\lambda+1}\right],$$

and the value at $\lambda = 0$ is taken to be the limit as $\lambda \to 0$ and λ is a real-valued parameter which is chosen by the user. Note that $\Phi^{(\lambda)}$ is expressed as the weighted geometric mean of the diversity index. The measure $\Phi^{(\lambda)}$ must lie between 0 and 1 since $0 \leq H_{ij}^{(\lambda)} \leq \left(2^{\lambda} - 1\right) / \left(\lambda 2^{\lambda}\right)$ for $i < j$. For any $\lambda \ (> -1)$, (i) $\Phi^{(\lambda)}$ takes the minimum value 0 if and only if there is a structure of the PS model in the table and (ii) $\Phi^{(\lambda)}$ takes the maximum value 1 if and only if the degree of departure from the PS model is the largest in the sense that $p_{ij}^{c} = 1$ (then $p_{ji}^{c} = 0$) or $p_{ji}^{c} = 1$ (then $p_{ij}^{c} = 0$) for all $i < j$. It is easily seen that the value of $\Phi^{(\lambda)}$ is less than or equal to the value of the measure of the S model proposed in [10, 11].

Assume that the observed frequencies $\{n_{ij}\}$ have a multinomial distribution. The sample version $\hat{\Phi}^{(\lambda)}$ of $\Phi^{(\lambda)}$ is given as $\Phi^{(\lambda)}$ with p_{ij} replaced by $\hat{p}_{ij}$, where $\hat{p}_{ij} = n_{ij}/n$ and $n = \sum\sum n_{ij}$. Using the delta method, $\sqrt{n}(\hat{\Phi}^{(\lambda)} - \Phi^{(\lambda)})$ has asymptotically (as $n \to \infty$) a normal distribution with mean zero and $\sigma^2\left[\Phi^{(\lambda)}\right]$(see [7] for the details of variances). Thus, when the sample size n is large, the estimated measure $\hat{\Phi}^{(\lambda)}$ is an asymptotically unbiased estimator of $\Phi^{(\lambda)}$. However, when the sample size n is not so large, the bias of $\hat{\Phi}^{(\lambda)}$ is large. Therefore, we consider improving the approximate unbiased estimator $\hat{\Phi}^{(\lambda)}$ by using the same method in [12].

§3. Improved approximate unbiased estimators

Let $\boldsymbol{p}$ be the $r^2 \times 1$ vector $\boldsymbol{p} = (p_{11}, \ldots, p_{1r}, p_{21}, \ldots, p_{2r}, \ldots, p_{r1}, p_{r2}, \ldots, p_{rr})^{\top}$ and let $\hat{\boldsymbol{p}}$ be the $r^2 \times 1$ vector in the similar way. $\boldsymbol{a}^{\top}$ denotes the transpose of $\boldsymbol{a}$. Using second Taylor expansion, the plug-in estimator $\hat{\Phi}^{(\lambda)}$ is expressed as

$$\hat{\Phi}^{(\lambda)} = \Phi^{(\lambda)} + \frac{\partial \Phi^{(\lambda)}}{\partial \boldsymbol{p}^\top}(\hat{\boldsymbol{p}} - \boldsymbol{p}) + \frac{1}{2}(\hat{\boldsymbol{p}} - \boldsymbol{p})^\top \frac{\partial^2 \Phi^{(\lambda)}}{\partial \boldsymbol{p} \partial \boldsymbol{p}^\top}(\hat{\boldsymbol{p}} - \boldsymbol{p}) + o_p(n^{-1}).$$

When the sample size n is large, $\mathbb{E}(\hat{\Phi}^{(\lambda)})$ is approximately equal to

$$\begin{aligned} \mathbb{E}(\hat{\Phi}^{(\lambda)}) &= \Phi^{(\lambda)} + \frac{1}{2}\mathrm{tr}\left(\left[\frac{\partial^2 \Phi^{(\lambda)}}{\partial \boldsymbol{p} \partial \boldsymbol{p}^\top}\right] \mathbb{E}((\hat{\boldsymbol{p}} - \boldsymbol{p})(\hat{\boldsymbol{p}} - \boldsymbol{p})^\top)\right) + o(n^{-1}) \\ &= \Phi^{(\lambda)} + \frac{1}{2n}\mathrm{tr}\left(\frac{\partial^2 \Phi^{(\lambda)}}{\partial \boldsymbol{p} \partial \boldsymbol{p}^\top}(D(\boldsymbol{p}) - \boldsymbol{p}\boldsymbol{p}^\top)\right) + o(n^{-1}), \end{aligned} \tag{3.1}$$

where $D(\boldsymbol{p})$ denotes the $r^2 \times r^2$ diagonal matrix with the ith element of $\boldsymbol{p}$ as the ith diagonal element. Thus it holds that

$$\mathbb{E}\left[\hat{\Phi}^{(\lambda)} - \frac{1}{2n}\mathrm{tr}\left(\left[\frac{\partial^2 \Phi^{(\lambda)}}{\partial \boldsymbol{p} \partial \boldsymbol{p}^\top}\right](D(\boldsymbol{p}) - \boldsymbol{p}\boldsymbol{p}^\top)\right)\right] = \Phi^{(\lambda)} + o(n^{-1}).$$

It is easy to know that it would approach $\Phi^{(\lambda)}$ faster than $\hat{\Phi}^{(\lambda)}$ as n tends to infinity. Therefore we now propose the improved approximate unbiased estimator $\hat{\Phi}^{*(\lambda)}$ of the true measure $\Phi^{(\lambda)}$ as follows:

$$\begin{aligned} \hat{\Phi}^{*(\lambda)} &= \hat{\Phi}^{(\lambda)} - \frac{1}{2n}\mathrm{tr}\left\{\widehat{\frac{\partial^2 \Phi^{(\lambda)}}{\partial \boldsymbol{p} \partial \boldsymbol{p}^\top}}(D(\hat{\boldsymbol{p}}) - \hat{\boldsymbol{p}}\hat{\boldsymbol{p}}^\top)\right\} \\ &= \hat{\Phi}^{(\lambda)} - \frac{1}{2n}\left\{\sum_{i \neq j} \hat{p}_{ij} \widehat{\frac{\partial^2 \Phi^{(\lambda)}}{\partial p_{ij} \partial p_{ij}}} - \sum_{(i,j)} \sum_{(k,l)} \hat{p}_{ij}\hat{p}_{kl} \widehat{\frac{\partial^2 \Phi^{(\lambda)}}{\partial p_{ij} \partial p_{kl}}}\right\}. \end{aligned}$$

Here $\widehat{\partial^2 \Phi^{(\lambda)}/\partial p_{ij}\partial p_{kl}}$ and $\widehat{\partial^2 \Phi^{(\lambda)}/\partial \boldsymbol{p}\partial \boldsymbol{p}^\top}$ are given as $\partial^2 \Phi^{(\lambda)}/\partial p_{ij}\partial p_{kl}$ and $\partial^2 \Phi^{(\lambda)}/\partial \boldsymbol{p}\partial \boldsymbol{p}^\top$ with p_{ij} replaced by $\hat{p}_{ij}$, respectively. The elements of the Hessian matrix $\partial^2 \Phi^{(\lambda)}/\partial \boldsymbol{p}\partial \boldsymbol{p}^\top$ are given in Appendix. Then we have the following theorem.

Theorem 3.1. *Let $\{\hat{p}_{ij}\}$ denote the maximum likelihood estimators for $i, j = 1, \ldots, r$. Then, as $n \to \infty$, it holds that*

$$\mathrm{Bias}\left(\hat{\Phi}^{*(\lambda)}\right) = \mathbb{E}\left[\hat{\Phi}^{*(\lambda)}\right] - \Phi^{(\lambda)} = o(n^{-1}).$$

$\hat{\Phi}^{*(\lambda)}$ tends to approximate to $\Phi^{(\lambda)}$ better than $\hat{\Phi}^{(\lambda)}$ when n is not so large. However, it does not always fall within the range $[0, 1]$ when the value of $\Phi^{(\lambda)}$

is close to 0 or 1. Therefore, in addition to the improved approximate unbiased estimator $\hat{\Phi}^{*(\lambda)}$, we also propose another improved approximate unbiased estimator whose value always falls within $[0,1]$. We consider that the monotone function $f:(0,1)\to\mathbb{R}$ be 2 times differentiable on $(0,1)$, for example, the logit function which is defined as

$$f(p)=\log\left(\frac{p}{1-p}\right).$$

The improved approximate unbiased estimator $\hat{\Phi}_f^{(\lambda)}$ using f is defined as

$$\hat{\Phi}_f^{(\lambda)}=f^{-1}\left(f(\hat{\Phi}^{(\lambda)})-\frac{1}{2n}f'(\hat{\Phi}^{(\lambda)})\mathrm{tr}\left\{\widehat{\frac{\partial^2\Phi^{(\lambda)}}{\partial\boldsymbol{p}\partial\boldsymbol{p}^\top}}(D(\hat{\boldsymbol{p}})-\hat{\boldsymbol{p}}\hat{\boldsymbol{p}}^\top)\right\}\right),$$

where f^{-1} and f' indicate the inverse function and the derivative function of f. It is clear that $\hat{\Phi}_f^{(\lambda)}$ always falls within the range of $[0,1]$ even when the value of $\Phi^{(\lambda)}$ is close to 0 or 1. Then it holds that the following theorem.

Theorem 3.2. *Let $\{\hat{p}_{ij}\}$ denote the maximum likelihood estimators for $i,j=1,\dots,r$. Then, as $n\to\infty$, it holds that*

$$\mathrm{Bias}\left(\hat{\Phi}_f^{(\lambda)}\right)=\mathbb{E}\left[\hat{\Phi}_f^{(\lambda)}\right]-\Phi^{(\lambda)}=o(n^{-1}).$$

Proof. Put $\boldsymbol{u}=\sqrt{n}(\hat{\boldsymbol{p}}-\boldsymbol{p})$ and

$$S=\frac{1}{2}f'\left(\hat{\Phi}^{(\lambda)}\right)\mathrm{tr}\left\{\widehat{\frac{\partial^2\Phi^{(\lambda)}}{\partial\boldsymbol{p}\partial\boldsymbol{p}^\top}}(D(\hat{\boldsymbol{p}})-\hat{\boldsymbol{p}}\hat{\boldsymbol{p}}^\top)\right\}. \tag{3.2}$$

By using Taylor expansion, it is calculated as follows:

$$\begin{aligned}\hat{\Phi}_f^{(\lambda)}&=f^{-1}\left(f(\hat{\Phi}^{(\lambda)})-\frac{1}{n}S\right)\\&=f^{-1}\left(f\left(\Phi^{(\lambda)}+\frac{1}{\sqrt{n}}\frac{\partial\Phi^{(\lambda)}}{\partial\boldsymbol{p}^\top}\boldsymbol{u}+\frac{1}{2n}\boldsymbol{u}^\top\frac{\partial^2\Phi^{(\lambda)}}{\partial\boldsymbol{p}\partial\boldsymbol{p}^\top}\boldsymbol{u}+o_p(n^{-1})\right)-\frac{1}{n}S\right)\\&=f^{-1}\Bigg(f(\Phi^{(\lambda)})+\frac{1}{\sqrt{n}}f'(\Phi^{(\lambda)})\frac{\partial\Phi^{(\lambda)}}{\partial\boldsymbol{p}^\top}\boldsymbol{u}+\frac{1}{2n}f'(\Phi^{(\lambda)})\boldsymbol{u}^\top\frac{\partial^2\Phi^{(\lambda)}}{\partial\boldsymbol{p}\partial\boldsymbol{p}^\top}\boldsymbol{u}\\&\qquad+\frac{1}{2n}f''(\Phi^{(\lambda)})\left(\frac{\partial\Phi^{(\lambda)}}{\partial\boldsymbol{p}^\top}\boldsymbol{u}\right)^2-\frac{1}{n}S+o_p(n^{-1})\Bigg)\\&=\Phi^{(\lambda)}+\frac{1}{\sqrt{n}}\frac{\partial\Phi^{(\lambda)}}{\partial\boldsymbol{p}^\top}\boldsymbol{u}+\frac{1}{2n}\boldsymbol{u}^\top\frac{\partial^2\Phi^{(\lambda)}}{\partial\boldsymbol{p}\partial\boldsymbol{p}^\top}\boldsymbol{u}\end{aligned}$$

$$
\begin{aligned}
&+ \frac{1}{2n}\frac{f''(\Phi^{(\lambda)})}{f'(\Phi^{(\lambda)})}\left(\frac{\partial\Phi^{(\lambda)}}{\partial\boldsymbol{p}^\top}\boldsymbol{u}\right)^2 - \frac{1}{n}\frac{1}{f'(\Phi^{(\lambda)})}S \\
&- \frac{1}{2n}\frac{f''(\Phi^{(\lambda)})}{\{f'(\Phi^{(\lambda)})\}^3}\left(f'(\Phi^{(\lambda)})\frac{\partial\Phi^{(\lambda)}}{\partial\boldsymbol{p}^\top}\boldsymbol{u}\right)^2 + o_p(n^{-1}) \\
= \Phi^{(\lambda)} &+ \frac{1}{\sqrt{n}}\frac{\partial\Phi^{(\lambda)}}{\partial\boldsymbol{p}^\top}\boldsymbol{u} + \frac{1}{2n}\boldsymbol{u}^\top\frac{\partial^2\Phi^{(\lambda)}}{\partial\boldsymbol{p}\partial\boldsymbol{p}^\top}\boldsymbol{u} - \frac{1}{n}\frac{1}{f'(\Phi^{(\lambda)})}S + o_p(n^{-1}).
\end{aligned}
$$

It is noted that the derivative functions of f^{-1} are given as

$$
\frac{\mathrm{d}}{\mathrm{d}x}f^{-1}(x) = \frac{1}{f'(f^{-1}(x))} \quad \text{and} \quad \frac{\mathrm{d}^2}{\mathrm{d}x^2}f^{-1}(x) = -\frac{f''(f^{-1}(x))}{\{f'(f^{-1}(x))\}^2}.
$$

Then it is given as

$$
\begin{aligned}
\mathbb{E}(\hat{\Phi}_f^{(\lambda)}) &= \mathbb{E}\left[\Phi^{(\lambda)} + \frac{1}{\sqrt{n}}\frac{\partial\Phi^{(\lambda)}}{\partial\boldsymbol{p}^\top}\boldsymbol{u} + \frac{1}{2n}\boldsymbol{u}^\top\frac{\partial^2\Phi^{(\lambda)}}{\partial\boldsymbol{p}\partial\boldsymbol{p}^\top}\boldsymbol{u} - \frac{1}{n}\frac{1}{f'(\Phi^{(\lambda)})}S\right] + o(1) \\
&= \Phi^{(\lambda)} + \frac{1}{2n}\mathrm{tr}\left\{\frac{\partial^2\Phi^{(\lambda)}}{\partial\boldsymbol{p}\partial\boldsymbol{p}^\top}\left(D(\boldsymbol{p}) - \boldsymbol{p}\boldsymbol{p}^\top\right)\right\} - \frac{1}{n}\frac{1}{f'(\Phi^{(\lambda)})}\mathbb{E}(S) + o(n^{-1}).
\end{aligned}
$$

From (3.2), it holds that

$$
\mathbb{E}(S) = \frac{1}{2}f'\left(\Phi^{(\lambda)}\right)\mathrm{tr}\left\{\frac{\partial^2\Phi^{(\lambda)}}{\partial\boldsymbol{p}\partial\boldsymbol{p}^\top}\left(D(\boldsymbol{p}) - \boldsymbol{p}\boldsymbol{p}^\top\right)\right\} + o(1).
$$

Therefore, we obtain as follows:

$$
\mathbb{E}(\hat{\Phi}_f^{(\lambda)}) = \Phi^{(\lambda)} + o(n^{-1}).
$$

□

From the above, the asymptotic bias of $\hat{\Phi}_f^{(\lambda)}$ improves over that of $\hat{\Phi}^{(\lambda)}$. The proposed estimator $\hat{\Phi}_f^{(\lambda)}$ tends to approximate to $\Phi^{(\lambda)}$ better than $\hat{\Phi}^{(\lambda)}$ when n is not so large.

§4. Simulation studies

In this section, we demonstrate the accuracy of two proposed estimators $\hat{\Phi}^{*(\lambda)}$ and $\hat{\Phi}_f^{(\lambda)}$ in the finite sample cases. Here we used the logit transformation as function f, that is, $f(x) = \log(x/(1-x))$. In this simulation, we consider artificial probability tables given in Table 1(a) to Table 1(f). Table 2 gives the value of $\Phi^{(\lambda)}$ for each λ for Table 1. For each λ, the true values of $\Phi^{(\lambda)}$ for Table 1(a) and Table 1(d) are close to 1, and the true values of $\Phi^{(\lambda)}$ for Table 1(c) and Table 1(f) are close to 0, and the true values of $\Phi^{(\lambda)}$ for Table 1(b) and Table 1(e) take around 0.5.

Using Monte Carlo simulation repeated 10000 times, we calculate the bias and the mean square error (MSE) of $\hat{\Phi}^{(\lambda)}$, $\hat{\Phi}^{*(\lambda)}$ and $\hat{\Phi}_f^{(\lambda)}$ from the observed frequencies of sample size $n = 500$, 1000, 3000, and 5000, which are obtained from the true probability distribution for Table 1(a) to Table 1(f). Table 3 to Table 8 give the bias values of $\hat{\Phi}^{(\lambda)}$, $\hat{\Phi}^{*(\lambda)}$ and $\hat{\Phi}_f^{(\lambda)}$ and the relative efficiency of the proposed estimators to $\hat{\Phi}^{(\lambda)}$ corresponding Table 1(a) to Table 1(f). Here $\mathrm{Bias}(\hat{\Phi})$ indicates the bias value of an estimator $\hat{\Phi}$, and the relative efficiency of an estimator $\hat{\Phi}$ to $\hat{\Phi}^{(\lambda)}$ is given by

$$\mathrm{e}(\hat{\Phi}) = \frac{\mathrm{MSE}(\hat{\Phi}^{(\lambda)})}{\mathrm{MSE}(\hat{\Phi})},$$

where $\mathrm{MSE}(\hat{\Phi})$ indicates MSE of an estimator $\hat{\Phi}$, that is, $\mathrm{MSE}(\hat{\Phi}) = \mathbb{E}\{(\hat{\Phi} - \Phi^{(\lambda)})^2\}$.

From Table 3 to Table 8, we see that the proposed estimators have a smaller bias than $\hat{\Phi}^{(\lambda)}$. This means that the proposed estimators approach to the true value faster than $\hat{\Phi}^{(\lambda)}$. In addition, it can be seen that the relative efficiency of proposed estimators are often greater than 1, which are more efficient than $\hat{\Phi}^{(\lambda)}$. From Table 8, when the true value of the measure is close to 0 and n is not so large, $\hat{\Phi}_f^{(\lambda)}$ has a larger bias than $\hat{\Phi}^{(\lambda)}$ and $\hat{\Phi}^{*(\lambda)}$. In addition, when the true value of the measure is close to 0 and n is not so large, the relative efficiency is less than 1. From these results, we can see that the proposed estimators become unstable when $\Phi^{(\lambda)}$ is close to 0 or 1. This is probably because the variance of $\hat{\Phi}^{(\lambda)}$ is small and $\hat{\Phi}^{(\lambda)}$ is a stable estimation. On the other hand, if the sample size is enough large, the proposed estimators improve the bias of $\hat{\Phi}^{(\lambda)}$, which are more efficient than $\hat{\Phi}^{(\lambda)}$.

§5. Concluding remarks

This paper proposed the improved approximate unbiased estimators $\hat{\Phi}^{*(\lambda)}$ and $\hat{\Phi}_f^{(\lambda)}$ of the true measure $\Phi^{(\lambda)}$. From the simulation studies, we conclude that if the sample size is enough large, the proposed estimators have a smaller bias than $\hat{\Phi}^{(\lambda)}$ and improve MSE. In other words, the proposed estimators converge to the true value $\Phi^{(\lambda)}$ faster than $\hat{\Phi}^{(\lambda)}$. However, the proposed estimators become unstable when $\Phi^{(\lambda)}$ is close to 0 or 1. Therefore, it is better to use the proposed estimators than $\hat{\Phi}^{(\lambda)}$ when the value of $\hat{\Phi}^{(\lambda)}$ is not close to 0 or 1, for example, it falls within the range $[0.2, 0.8]$. On the other hand, it is better to use a plug-in estimator if $\hat{\Phi}^{(\lambda)}$ is close to 0 or 1 and n is not so large.

Acknowledgments

The authors would like to thank the referee and the editor for many comments and suggestions. This work was supported by JSPS Grant-in-Aid for Scientific Research (C) Number JP20K03756, JSPS Grant-in-Aid for Scientific Research (C) Number JP18K03425 and JSPS Grant-in-Aid for Early-Career Scientists JP19K14600.

§A. The elements of the second derivative of $\Phi^{(\lambda)}$

For $\lambda > -1$, the elements of $\partial^2\Phi^{(\lambda)}/\partial \boldsymbol{p}\partial \boldsymbol{p}^\top$ ($r^2 \times r^2$ matrix) are given as follows:

$$\begin{aligned}
\frac{\partial^2\Phi^{(\lambda)}}{\partial p_{ij}\partial p_{ij}} &= \frac{\Phi^{(\lambda)}}{\delta^2}\left\{K_{ij}^{(\lambda)}K_{ij}^{(\lambda)} - 2K_{ij}^{(\lambda)} + \frac{\delta p_{ji}^c}{(p_{ij}+p_{ji})}J_{ij}^{(\lambda)}\right\} && (i \neq j),\\
\frac{\partial^2\Phi^{(\lambda)}}{\partial p_{ij}\partial p_{ji}} &= \frac{\Phi^{(\lambda)}}{\delta^2}\left\{K_{ij}^{(\lambda)}K_{ji}^{(\lambda)} - \left(K_{ij}^{(\lambda)} + K_{ji}^{(\lambda)}\right) - \frac{\delta p_{ij}^c}{(p_{ij}+p_{ji})}J_{ij}^{(\lambda)}\right\} && (i \neq j),\\
\frac{\partial^2\Phi^{(\lambda)}}{\partial p_{ij}\partial p_{kl}} &= 0 && (i = j \text{ or } k = l),\\
\frac{\partial^2\Phi^{(\lambda)}}{\partial p_{ij}\partial p_{kl}} &= \frac{\Phi^{(\lambda)}}{\delta^2}\left\{K_{ij}^{(\lambda)}K_{kl}^{(\lambda)} - \left(K_{ij}^{(\lambda)} + K_{kl}^{(\lambda)}\right)\right\} && (\text{otherwise}),
\end{aligned}$$

where

$$\begin{aligned}
K_{ij}^{(\lambda)} &= \log(w_{ij}^{(\lambda)}) + \frac{\xi_{ij}^{(\lambda)}}{w_{ij}^{(\lambda)}} - \log(\Phi^{(\lambda)}),\\
J_{ij}^{(\lambda)} &= \frac{2^\lambda(\lambda+1)}{(2^\lambda-1)w_{ij}^{(\lambda)}}\left\{\lambda p_{ji}^c\left((p_{ij}^c)^{\lambda-1} + (p_{ji}^c)^{\lambda-1}\right) - \left((p_{ij}^c)^{\lambda} - (p_{ji}^c)^{\lambda}\right)\frac{\xi_{ij}^{(\lambda)}}{w_{ij}^{(\lambda)}}\right\},\\
w_{ij}^{(\lambda)} &= 1 - \frac{2^\lambda}{2^\lambda-1}\left\{1 - (p_{ij}^c)^{\lambda+1} - (p_{ji}^c)^{\lambda+1}\right\},\\
\xi_{ij}^{(\lambda)} &= \frac{2^\lambda}{2^\lambda-1}(\lambda+1)p_{ji}^c\left\{(p_{ij}^c)^{\lambda} - (p_{ji}^c)^{\lambda}\right\}.
\end{aligned}$$

In particular, when $\lambda = 0$, it is given as follows:

$$K_{ij}^{(0)} = \lim_{\lambda\to 0} K_{ij}^{(\lambda)} = \log(w_{ij}^{(0)}) + \frac{\xi_{ij}^{(0)}}{w_{ij}^{(0)}} - \log(\Phi^{(0)}),$$

$$J_{ij}^{(0)} = \lim_{\lambda \to 0} J_{ij}^{(\lambda)} = \frac{p_{ji}^c \left[(p_{ij}^c)^{-1} + (p_{ji}^c)^{-1} \right] - \left[\log(p_{ij}^c) - \log(p_{ji}^c) \right] (\xi_{ij}^{(0)} / w_{ij}^{(0)})}{w_{ij}^{(0)} \log 2},$$

$$w_{ij}^{(0)} = 1 - \frac{1}{\log 2} \left(-p_{ij}^c \log p_{ij}^c - p_{ji}^c \log p_{ji}^c \right),$$

$$\xi_{ij}^{(0)} = \frac{1}{\log 2} p_{ji}^c \left(\log p_{ij}^c - \log p_{ji}^c \right).$$

References

[1] A. Agresti, *Categorical Data Analysis*, 2nd ed. (2002), Wiley, New York.

[2] Y. M. M. Bishop, S. E. Fienberg and P. W. Holland, *Discrete Multivariate Analysis: Theory and Practice*, (1975), MIT Press, Cambridge, MA.

[3] A. H. Bowker, A test for symmetry in contingency tables, *Journal of American Statistical Association* **43** (1948), 572–574.

[4] H. Caussinus, Contribution à l'analyse statistique des tableaux de corrélation, *Annales de la Faculté des Sciences de Toulouse*, **29** (1965), 77–182.

[5] P. McCullagh, A class of parametric models for the analysis of square contingency tables with ordered categories, *Biometrika*, **65** (1978), 413–418.

[6] A. Stuart, A test for homogeneity of the marginal distributions in a two-way classification. *Biometrika*, **42** (1955), 412–416.

[7] Y. Saigusa, K. Tahata and S. Tomizawa, Measure of departure from partial symmetry for square contingency tables, *Journal of Mathematics and Statistics* **12** (2016), 152–156.

[8] K. Tahata, H. Tanaka and S. Tomizawa, Refined estimators of measures for marginal homogeneity in square contingency tables. *International Journal of Pure and Applied Mathematics* **90** (2014), 501–513.

[9] K. Tahata, R. Tomisato and S. Tomizawa, An improved approximate unbiased estimator of log-odds for 2×2 contingency tables. *Advances and Applications in Statistics* **9** (2008), 1–12.

[10] S. Tomizawa, Two kinds of measures of departure from symmetry in square contingency tables having nominal categories, *Statistica Sinica* **4** (1994), 325–334.

[11] S. Tomizawa, T. Seo and H. Yamamoto, Power-divergence-type measure of departure from symmetry for square contingency tables that have nominal categories, *Journal of Applied Statistics* **25** (1998), 387–398.

[12] S. Tomizawa, N. Miyamoto and N. Ohba, Improved approximate unbiased estimators of measures of asymmetry for square contingency tables, *Advances and Applications in Statistics* **7** (2007), 47–63.

Table 1: Artificial probability table.

(a)

	(1)	(2)	(3)
(1)	0.000	0.011	0.011
(2)	0.322	0.000	0.011
(3)	0.322	0.323	0.000

(b)

	(1)	(2)	(3)
(1)	0.000	0.037	0.037
(2)	0.297	0.000	0.037
(3)	0.296	0.296	0.000

(c)

	(1)	(2)	(3)
(1)	0.000	0.095	0.095
(2)	0.239	0.000	0.095
(3)	0.238	0.238	0.000

(d)

	(1)	(2)	(3)	(4)
(1)	0.000	0.004	0.005	0.005
(2)	0.161	0.000	0.005	0.005
(3)	0.162	0.162	0.000	0.005
(4)	0.162	0.162	0.162	0.000

(e)

	(1)	(2)	(3)	(4)
(1)	0.000	0.004	0.005	0.005
(2)	0.161	0.000	0.005	0.005
(3)	0.162	0.162	0.000	0.005
(4)	0.162	0.162	0.162	0.000

(f)

	(1)	(2)	(3)	(4)
(1)	0.000	0.055	0.056	0.056
(2)	0.111	0.000	0.056	0.056
(3)	0.111	0.110	0.000	0.056
(4)	0.111	0.111	0.111	0.000

Table 2: The values of $\Phi^{(\lambda)}$ for each λ corresponding to artificial probability tables in Table 1.

λ	0.0	0.5	1.0	1.5
Table 1(a)	0.791	0.853	0.872	0.876
Table 1(b)	0.497	0.574	0.605	0.612
Table 1(c)	0.138	0.169	0.185	0.189
Table 1(d)	0.811	0.869	0.887	0.890
Table 1(e)	0.491	0.568	0.599	0.606
Table 1(f)	0.080	0.099	0.109	0.111

Table 3: The bias values of the estimators multiplied by 100, and the ratios of MSE of the proposed estimators to $\hat{\Phi}^{(\lambda)}$ for Table 1(a).

n	λ	Bias($\hat{\Phi}^{(\lambda)}$)	Bias($\hat{\Phi}^{*(\lambda)}$)	Bias($\hat{\Phi}_f^{(\lambda)}$)	e($\hat{\Phi}^{*(\lambda)}$)	e($\hat{\Phi}_f^{(\lambda)}$)
500	0.0	0.345	0.103	0.101	1.018	1.018
	0.5	0.064	0.032	0.032	1.003	1.003
	1.0	−0.029	−0.003	−0.003	1.004	1.004
	1.5	−0.030	0.003	0.003	1.005	1.005
1000	0.0	0.080	−0.041	−0.042	1.007	1.007
	0.5	0.005	−0.012	−0.012	1.001	1.001
	1.0	−0.032	−0.019	−0.019	1.002	1.002
	1.5	0.007	0.024	0.024	1.002	1.002
3000	0.0	0.044	0.003	0.003	1.003	1.003
	0.5	0.000	−0.006	−0.006	1.000	1.000
	1.0	−0.010	−0.006	−0.006	1.001	1.001
	1.5	−0.015	−0.009	−0.009	1.001	1.001
5000	0.0	0.030	0.006	0.006	1.002	1.002
	0.5	0.008	0.005	0.005	1.000	1.000
	1.0	−0.005	−0.002	−0.002	1.000	1.000
	1.5	−0.018	−0.014	−0.014	1.001	1.001

Table 4: The bias values of the estimators multiplied by 100, and the ratios of MSE of the proposed estimators to $\hat{\Phi}^{(\lambda)}$ for Table 1(b).

n	λ	Bias($\hat{\Phi}^{(\lambda)}$)	Bias($\hat{\Phi}^{*(\lambda)}$)	Bias($\hat{\Phi}_f^{(\lambda)}$)	e($\hat{\Phi}^{*(\lambda)}$)	e($\hat{\Phi}_f^{(\lambda)}$)
500	0.0	0.067	−0.008	−0.008	1.010	1.010
	0.5	−0.041	−0.013	−0.013	1.005	1.005
	1.0	−0.095	−0.015	−0.015	1.005	1.005
	1.5	−0.024	0.067	0.067	1.004	1.004
1000	0.0	0.070	0.032	0.032	1.005	1.005
	0.5	−0.060	−0.046	−0.046	1.003	1.003
	1.0	−0.011	0.028	0.028	1.002	1.002
	1.5	−0.026	0.019	0.019	1.002	1.002
3000	0.0	0.036	0.023	0.023	1.002	1.002
	0.5	0.012	0.017	0.017	1.001	1.001
	1.0	−0.026	−0.012	−0.012	1.001	1.001
	1.5	−0.017	−0.002	−0.002	1.001	1.001
5000	0.0	−0.019	−0.026	−0.026	1.001	1.001
	0.5	0.009	0.012	0.012	1.000	1.000
	1.0	−0.017	−0.009	−0.009	1.001	1.001
	1.5	−0.004	0.005	0.005	1.000	1.000

Table 5: The bias values of the estimators multiplied by 100, and the ratios of MSE of the proposed estimators to $\hat{\Phi}^{(\lambda)}$ for Table 1(c).

n	λ	Bias($\hat{\Phi}^{(\lambda)}$)	Bias($\hat{\Phi}^{*(\lambda)}$)	Bias($\hat{\Phi}_f^{(\lambda)}$)	e($\hat{\Phi}^{*(\lambda)}$)	e($\hat{\Phi}_f^{(\lambda)}$)
500	0.0	−0.119	−0.026	−0.025	1.012	1.012
	0.5	−0.167	−0.025	−0.025	1.010	1.010
	1.0	−0.106	0.064	0.065	1.006	1.007
	1.5	−0.131	0.048	0.049	1.006	1.007
1000	0.0	−0.039	0.006	0.006	1.005	1.005
	0.5	−0.071	−0.002	−0.002	1.004	1.004
	1.0	−0.095	−0.012	−0.012	1.004	1.004
	1.5	−0.053	0.034	0.034	1.002	1.002
3000	0.0	−0.009	0.006	0.006	1.001	1.001
	0.5	−0.021	0.001	0.001	1.001	1.001
	1.0	−0.027	0.000	0.000	1.001	1.001
	1.5	−0.006	0.023	0.023	1.000	1.000
5000	0.0	−0.015	−0.006	−0.006	1.001	1.001
	0.5	−0.041	−0.027	−0.027	1.001	1.001
	1.0	−0.012	0.005	0.005	1.001	1.001
	1.5	−0.021	−0.004	−0.004	1.001	1.001

Table 6: The bias values of the estimators multiplied by 100, and the ratios of MSE of the proposed estimators to $\hat{\Phi}^{(\lambda)}$ for Table 1(d).

n	λ	Bias($\hat{\Phi}^{(\lambda)}$)	Bias($\hat{\Phi}^{*(\lambda)}$)	Bias($\hat{\Phi}_f^{(\lambda)}$)	e($\hat{\Phi}^{*(\lambda)}$)	e($\hat{\Phi}_f^{(\lambda)}$)
500	0.0	0.502	0.274	0.272	0.976	0.976
	0.5	0.028	0.027	0.027	1.006	1.006
	1.0	−0.063	0.029	0.029	1.017	1.017
	1.5	−0.165	−0.062	−0.063	1.021	1.021
1000	0.0	0.201	−0.002	−0.003	1.014	1.014
	0.5	0.032	0.027	0.027	1.006	1.006
	1.0	−0.057	−0.011	−0.011	1.009	1.009
	1.5	−0.072	−0.021	−0.021	1.010	1.010
3000	0.0	0.077	0.006	0.006	1.008	1.008
	0.5	0.031	0.028	0.028	1.002	1.002
	1.0	−0.013	0.002	0.002	1.003	1.003
	1.5	−0.016	0.001	0.001	1.003	1.003
5000	0.0	0.054	0.012	0.012	1.005	1.005
	0.5	−0.001	−0.003	−0.003	1.001	1.001
	1.0	−0.013	−0.004	−0.004	1.002	1.002
	1.5	−0.012	−0.002	−0.002	1.002	1.002

Table 7: The bias values of the estimators multiplied by 100, and the ratios of MSE of the proposed estimators to $\hat{\Phi}^{(\lambda)}$ for Table 1(e).

n	λ	Bias($\hat{\Phi}^{(\lambda)}$)	Bias($\hat{\Phi}^{*(\lambda)}$)	Bias($\hat{\Phi}_f^{(\lambda)}$)	e($\hat{\Phi}^{*(\lambda)}$)	e($\hat{\Phi}_f^{(\lambda)}$)
500	0.0	−0.024	0.018	0.018	1.026	1.026
	0.5	−0.225	0.016	0.015	1.021	1.021
	1.0	−0.393	−0.059	−0.060	1.025	1.025
	1.5	−0.402	−0.048	−0.049	1.026	1.026
1000	0.0	0.010	0.029	0.029	1.013	1.013
	0.5	−0.120	−0.003	−0.003	1.010	1.010
	1.0	−0.168	−0.005	−0.005	1.011	1.011
	1.5	−0.204	−0.031	−0.031	1.013	1.013
3000	0.0	−0.022	−0.016	−0.016	1.004	1.004
	0.5	−0.023	0.015	0.015	1.003	1.003
	1.0	−0.093	−0.039	−0.039	1.005	1.005
	1.5	−0.034	0.023	0.023	1.003	1.003
5000	0.0	−0.005	−0.002	−0.002	1.003	1.003
	0.5	−0.050	−0.027	−0.027	1.003	1.003
	1.0	−0.068	−0.036	−0.036	1.003	1.003
	1.5	−0.046	−0.012	−0.012	1.003	1.003

Table 8: The bias values of the estimators multiplied by 100, and the ratios of MSE of the proposed estimators to $\hat{\Phi}^{(\lambda)}$ for Table 1(f).

n	λ	Bias($\hat{\Phi}^{(\lambda)}$)	Bias($\hat{\Phi}^{*(\lambda)}$)	Bias($\hat{\Phi}_f^{(\lambda)}$)	e($\hat{\Phi}^{*(\lambda)}$)	e($\hat{\Phi}_f^{(\lambda)}$)
500	0.0	−0.598	0.277	0.832	0.831	0.098
	0.5	−0.724	0.349	0.846	0.921	0.174
	1.0	−0.763	0.457	1.004	0.802	0.173
	1.5	−0.893	0.369	0.953	0.852	0.169
1000	0.0	−0.265	0.024	0.041	1.063	0.846
	0.5	−0.331	0.032	0.039	1.063	1.066
	1.0	−0.371	0.037	0.056	1.061	0.875
	1.5	−0.364	0.051	0.059	1.059	1.062
3000	0.0	−0.092	−0.006	−0.005	1.016	1.016
	0.5	−0.113	−0.003	−0.002	1.016	1.016
	1.0	−0.121	0.002	0.003	1.015	1.015
	1.5	−0.123	0.003	0.004	1.015	1.016
5000	0.0	−0.044	0.006	0.006	1.007	1.007
	0.5	−0.059	0.006	0.006	1.007	1.007
	1.0	−0.084	−0.011	−0.011	1.011	1.011
	1.5	−0.069	0.006	0.006	1.008	1.008

Tomoyuki Nakagawa
Department of Information Sciences, Tokyo University of Science
Noda City, Chiba 278-8510, Japan
E-mail: t_nakagawa@rs.tus.ac.jp

Ryoma Namba
Department of Information Sciences, Tokyo University of Science
Noda City, Chiba 278-8510, Japan

Kiyotaka Iki
Department of Economics, Nihon University
Chiyoda-ku, Tokyo 101-8360, Japan

Sadao Tomizawa
Department of Information Sciences, Tokyo University of Science
Noda City, Chiba 278-8510, Japan