symmetry

Article

Diagonals–Parameter Symmetry Model and Its Property for Square Contingency Tables with Ordinal Categories

Kouji Tahata * and **Kohei Matsuda**

Department of Information Sciences, Tokyo University of Science, Chiba 278-8510, Japan
* Correspondence: kouji_tahata@rs.tus.ac.jp

Abstract: The diagonals–parameter symmetry (DPS) model is a proposed method for analyzing square contingency tables with ordinal categories. Previously, it was stated that the generalized DPS (DPS[f]) model was equivalent to the DPS model for any function f, but the proof was not provided. This paper presents the derivation of the DPS[f] model and the proof of the relationship between the two models. The findings offer various interpretations of the DPS model. Additionally, a new model is considered, and it is shown that the proposed model and the DPS[f] model are separable.

Keywords: conditional symmetry; f-divergence; global symmetry; partial global symmetry

check for updates

Citation: Tahata, K.; Matsuda, K. Diagonals–Parameter Symmetry Model and Its Property for Square Contingency Tables with Ordinal Categories. *Symmetry* **2024**, *16*, 768. https://doi.org/10.3390/sym16060768

Academic Editors: Dušan Nikezić and Vesna Borka Jovanović

Received: 2 May 2024
Revised: 11 June 2024
Accepted: 14 June 2024
Published: 19 June 2024

1. Introduction

A contingency table with identical categories for rows and columns can be produced when a categorical variable is repeatedly measured. Observations in this type of table tend to concentrate on the cells along the main diagonal. Our research focuses on applying symmetry instead of assuming independence between row and column categories. Several studies have addressed symmetry issues, such as [1–9].

Let X and Y represent the row and column variables for an $r \times r$ contingency table with ordinal categories. Additionally, let π_{ij} represent the probability of an observation falling into the (i,j)th cell, where $i = 1,\dots,r$ and $j = 1,\dots,r$. The diagonals–parameter symmetry (DPS) model proposed by Goodman [10] is defined as follows.

$$\pi_{ij} = \begin{cases} d_k \psi_{ij} & (i < j), \\ \psi_{ij} & (i \geq j), \end{cases} \tag{1}$$

where $\psi_{ij} = \psi_{ji}$ and $k = j - i$. The parameter d_k in the DPS model represents the odds of an observation falling into cells (i,j) where $j - i = k$ and $i < j$, rather than cells (j,i) for $k = 1,\dots,r-1$. Moreover, the ratio between π_{ij} and π_{ji} can be expressed as the constant d_k for $j - i = k$ and $i < j$. This ratio depends solely on the distance from the main diagonal cells.

When $d_1 = d_2 = \cdots = d_{r-1} = 1$ in Equation (1), the DPS model reduces to the symmetry (S) model proposed by Bowker [1]. When d_k is independent of i and j in Equation (1), with $d_1 = \cdots = d_{r-1}$, the DPS model reduces to the conditional symmetry (CS) model proposed by McCullagh [11].

Using the f-divergence, Kateri and Papaioannou [2] proposed the generalized DPS (DPS[f]) model, defined as

$$\pi_{ij} = \pi_{ij}^{S} F^{-1}(\Delta_k + \zeta_{ij}) \quad (i = 1,\dots,r;\ j = 1,\dots,r), \tag{2}$$

where $k = i - j$, $\pi_{ij}^{S} = (\pi_{ij} + \pi_{ji})/2$, $\zeta_{ij} = \zeta_{ji}$ and $\Delta_k + \Delta_{-k} = 0$. It should be noted that the function f is twice-differentiable and strictly convex. Additionally, $F(t) = f'(t)$, $f(1) = 0$, $f(0) = \lim_{t\to 0} f(t)$, $0 \cdot f(0/0) = 0$, and $0 \cdot f(a/0) = a \lim_{t\to\infty}[f(t)/t]$. The model derivation is not included in their paper. They did mention that the DPS model is the closest to symmetry regarding the Kullback–Leibler distance under some conditions and that the DPS[f] model is equivalent to the DPS model. In this study, we will derive the DPS[f] model and provide proof of the relation between the two models. We can obtain various interpretations of the DPS model from the result. We discuss the necessary and sufficient condition for the S model and the property between test statistics for goodness of fit.

The paper is organized as follows: Section 2 derives Equation (2) and interprets the model from an information theory viewpoint. The proof is given that the DPS[f] model is equivalent to the DPS model regardless of the function f. Section 3 considers a new model and proves that the proposed model and the DPS[f] model are separable. A numerical example is provided in Section 4. Finally, Section 5 summarizes the paper.

2. Properties of the DPS[f] Model

Kateri and Papaioannou [2] noted that the DPS[f] model is the closest model to the S model in terms of the f-divergence under the conditions where $\sum\sum_{j-i=k}\pi_{ij}$ (and $\sum\sum_{i-j=k}\pi_{ij}$) for k=$1,\ldots,r-1$ and the sums $\pi_{ij}+\pi_{ji}$ for $i,j=1,\ldots,r$ are given. Similar research has been conducted in, for example, Ireland et al. [12], Kateri and Agresti [3], and Tahata [5]. This section derives the DPS[f] model and describes its properties.

We can obtain the following theorem, although the proof of Theorem 1 is given in Appendix A.1.

Theorem 1. *In the class of models with given $\sum\sum_{i-j=k}\ \pi_{ij}$, $k\neq 0$, and $\pi_{ij}+\pi_{ji}$ $(i=1,\ldots,r;\ j=1,\ldots,r)$, the model*

$$\pi_{ij}=\pi_{ij}^{S}F^{-1}(\Delta_k+\zeta_{ij})\quad (i=1,\ldots,r;\ j=1,\ldots,r)$$

with $k=i-j$, $\zeta_{ij}=\zeta_{ji}$ and $\Delta_k+\Delta_{-k}=0$, is the model closest to the complete symmetry model in terms of the f-divergence.

The DPS[f] model can be expressed as

$$F(2\pi_{ij}^{c})=\begin{cases}\gamma_{ij}+a_k & (i<j),\\ \gamma_{ij} & (i\geq j),\end{cases}\tag{3}$$

where $k=j-i$, $\gamma_{ij}=\gamma_{ji}$ and $\pi_{ij}^{c}=\pi_{ij}/(\pi_{ij}+\pi_{ji})$. It should be noted that π_{ij}^{c} represents the conditional probability of an observation falling in the (i,j) cell, given that it falls in either the (i,j) cell or the (j,i) cell. Namely, the DPS[f] model indicates that

$$F(2\pi_{ij}^{c})-F(2\pi_{ji}^{c})=a_k\quad (i<j).\tag{4}$$

When $a_1=\cdots=a_{r-1}=0$, the DPS[f] model is reduced to the S model.

If $f(x)=x\log(x)$, $x>0$, then the f-divergence is reduced to the KL divergence. When we set $f(x)=x\log(x)$, Equation (3) is reduced to

$$\pi_{ij}=\begin{cases}\pi_{ij}^{S}\exp(\gamma_{ij}+a_k-1) & (i<j),\\ \pi_{ij}^{S}\exp(\gamma_{ij}-1) & (i\geq j),\end{cases}$$

where $k=j-i$ and $\gamma_{ij}=\gamma_{ji}$. We shall refer to this model as the DPS$_{\text{KL}}$ model. Under the DPS$_{\text{KL}}$ model, the ratios of π_{ij} and π_{ji} for $i<j$ are expressed as

$$\frac{\pi_{ij}}{\pi_{ji}}=d_k^{\text{KL}}\quad (i<j),\tag{5}$$

where $d_k^{\text{KL}}=\exp(a_k)$ and $k=j-i$. Since Equation (5) indicates that the ratio of π_{ij} and π_{ji} depends on the distance of $k=j-i$, the DPS$_{\text{KL}}$ model is equivalent to the DPS model proposed by Goodman [10]. Namely, the DPS model is the closest model to the S model in terms of the KL divergence under the conditions where $\sum\sum_{i-j=k} \pi_{ij}$, $k \neq 0$, and the sums $\pi_{ij} + \pi_{ji}$ for $i = 1, \ldots, r;\ j = 1, \ldots, r$ are given. This is a special case of Theorem 1.

If $f(x) = -\log(x)$, $x > 0$, then the f-divergence is reduced to the reverse KL divergence. Then, the DPS[f] model is reduced to

$$\pi_{ij} = \begin{cases} \pi^S_{ij}\left(-\dfrac{1}{\gamma_{ij} + a_k}\right) & (i < j), \\ \pi^S_{ij}\left(-\dfrac{1}{\gamma_{ij}}\right) & (i \geq j), \end{cases}$$

where $k = j - i$ and $\gamma_{ij} = \gamma_{ji}$. We shall refer to this model as the DPS_{RKL} model. This model is the closest to the S model when the divergence is measured by the reverse KL divergence and can be expressed as

$$\frac{1}{\pi^c_{ij}} - \frac{1}{\pi^c_{ji}} = d^{\text{RKL}}_k \quad (i < j),$$

where $d^{\text{RKL}}_k = -2a_k$ and $k = j - i$. This model indicates that the difference between inverse probabilities $1/\pi^c_{ij}$ and $1/\pi^c_{ji}$ depends on the distance of $k = j - i$.

If $f(x) = (1 - x)^2$, then the f-divergence is reduced to the χ^2-divergence (Pearsonian distance). Then, the DPS[f] model is reduced to

$$\pi_{ij} = \begin{cases} \pi^S_{ij}\left(\dfrac{\gamma_{ij} + a_k}{2} + 1\right) & (i < j), \\ \pi^S_{ij}\left(\dfrac{\gamma_{ij}}{2} + 1\right) & (i \geq j), \end{cases}$$

where $k = j - i$ and $\gamma_{ij} = \gamma_{ji}$. We shall refer to this model as the DPS_{P} model. This model is the closest to the S model when the divergence is measured by the χ^2-divergence and can be expressed as

$$\pi^c_{ij} - \pi^c_{ji} = d^{\text{P}}_k \quad (i < j),$$

where $d^{\text{P}}_k = a_k/4$ and $k = j - i$. This model indicates that the difference between π^c_{ij} and π^c_{ji} depends on the distance of $k = j - i$.

Moreover, if $f(x) = (\lambda(\lambda + 1))^{-1}(x^{\lambda+1} - x)$, $x > 0$, where λ is a real-valued parameter, then the f-divergence is reduced to the power-divergence [13]. Then, the DPS[f] model is reduced to

$$\pi_{ij} = \begin{cases} \pi^S_{ij}\left(\lambda(\gamma_{ij} + a_k) + \dfrac{1}{\lambda + 1}\right)^{\frac{1}{\lambda}} & (i < j), \\ \pi^S_{ij}\left(\lambda\gamma_{ij} + \dfrac{1}{\lambda + 1}\right)^{\frac{1}{\lambda}} & (i \geq j), \end{cases}$$

where $k = j - i$ and $\gamma_{ij} = \gamma_{ji}$. We shall refer to this model as the $\text{DPS}_{\text{PD}(\lambda)}$ model. This model is the closest to the S model when the power-divergence measures the divergence and can be expressed as

$$(\pi^c_{ij})^\lambda - (\pi^c_{ji})^\lambda = d^{\text{PD}(\lambda)}_k \quad (i < j),$$

where $d^{\text{PD}(\lambda)}_k = (\lambda a_k)/2^\lambda$ and $k = j - i$. This model indicates that the difference between the symmetric conditional probabilities to the power of λ depends on the distance of $k = j - i$. When we apply the $\text{DPS}_{\text{PD}(\lambda)}$ model, we should set the value of λ.

Kateri and Papaioannou [2] reported that the DPS[f] model is equivalent to the DPS model regardless of f. That is, all the models described above (i.e., $\mathrm{DPS_{KL}}$, $\mathrm{DPS_{RKL}}$, $\mathrm{DPS_{P}}$, and $\mathrm{DPS_{PD(\lambda)}}$) are equivalent to the DPS model surprisingly. However, the proof was not given. We prove the following theorem.

Theorem 2. *The DPS[f] model is equivalent to the DPS model regardless of f.*

The poof is given in Appendix A.2. Theorem 2 states that the DPS model holds if and only if the DPS[f] model holds. If the DPS model fits the given dataset, we obtain various interpretations for the data.

When $a_1 = \cdots = a_{r-1}$, the DPS[f] model is reduced to the conditional symmetry model based on the f-divergence (CS[f]) model. The CS[f] model is described previously Kateri and Papaioannou [2]. Additionally, Fujisawa and Tahata [14] proposed the generalization of the CS[f] model. Similarly, when $d_1 = \cdots = d_{r-1}$, the DPS model is reduced to the CS model proposed by McCullagh [11]. The CS[f] model is equivalent to the CS model regardless of f (Kateri and Papaioannou [2]). Hence, Theorem 2 leads to the following result.

Corollary 1. *The CS[f] model is equivalent to the CS model regardless of f.*

3. Equivalence Conditions for Symmetry

Here, the equivalence conditions of the S model are discussed. If the S model holds, then the DPS[f] model with $a_1 = \cdots = a_{r-1} = 0$ holds. Conversely, if the DPS[f] model holds, then the S model does not hold generally. Therefore, we are interested in considering an additional condition to obtain the S model when the DPS[f] model holds. Other studies have discussed such conditions; see Read [15] and Tahata et al. [16].

We consider the distance global symmetry (DGS) model defined as

$$\delta_k^U = \delta_k^L \quad (k = 1, \ldots, r-1),$$

where $\delta_k^U = \sum\sum_{j-i=k} \pi_{ij}$, $\delta_k^L = \sum\sum_{i-j=k} \pi_{ij}$. For $k = 1, \ldots, r-1$, this model indicates that the sum of probabilities which are apart distance $k = j - i$ from main diagonal cells is equal to the sum of probabilities which are apart distance $k = i - j$ from main diagonal cells. We obtain the following theorem. (The proof is given in Appendix A.3.)

Theorem 3. *The S model holds if and only if both the DPS[f] and DGS models hold.*

Next, we consider the global symmetry (GS) model, which is defined as

$$\sum_{i<j}\sum \pi_{ij} = \sum_{i<j}\sum \pi_{ji}.$$

It should be noted that the DGS model implies the GS model. Read [15] noted that the S model holds if and only if both the CS and GS models hold. Fujisawa and Tahata [14] proved that the S model holds if and only if the CS[f] and GS models hold. These statements are the same as those from Corollary 1. In addition, a refined estimator for measures associated with the S, CS, and GS models was introduced by [17]. The result has a significant connection to decomposing the S model and separating the goodness-of-fit test statistic of the S model. According to Corollary 1, the refined estimator for the measure of CS can be utilized to gauge the extent of deviation from the CS[f] model.

This section proves the separation of the test statistics for the S model into those for the DPS[f] model and the DGS model. Consider a square contingency table of size $r \times r$ where n_{ij} denotes the observed frequency in the cell located at the (i, j) position. Assume this contingency table adheres to a multinomial distribution. In this context, let m_{ij} represent the expected frequency in the (i, j) cell, and $\hat{m}_{ij}$ be its corresponding maximum likelihood estimate under a specified model. To test each model's goodness of fit, we can employ the likelihood ratio chi-square statistic, denoted by $G^2(M)$. This statistic is computed using the following formula:

$$G^2(M) = 2\sum_{i=1}^{r}\sum_{j=1}^{r} n_{ij}\log\left(\frac{n_{ij}}{\hat{m}_{ij}}\right).$$

This statistic follows a chi-square distribution with the corresponding degrees of freedom (df).

It is supposed that model M_3 holds if and only if both models M_1 and M_2 hold. In this case, if the analyst has found hypothesis M3 unacceptable, their attention will move to examining components M1 and M2. For these three models, Aitchison [18] discussed the properties of the Wald test statistics, and Darroch and Silvey [19] described the properties of the likelihood ratio chi-square statistics. Assume that the following equivalence holds:

$$T(\mathrm{M}_3) = T(\mathrm{M}_1) + T(\mathrm{M}_2), \tag{6}$$

where T is the goodness of fit test statistic and the number of df for M_3 is equal to the sum of numbers of df for M_1 and M_2. If both M_1 and M_2 are accepted with a high probability (at the α significance level), then M_3 is accepted. However, when (6) does not hold, an incompatible situation where both M_1 and M_2 are accepted with a high probability but M_3 is rejected may arise. In fact, Darroch and Silvey [19] showed such an interesting example. The partitions of chi-squared test statistics are also discussed in, for example, [20,21].

From Theorem 3, the S model holds if and only if the DPS[f] model and the DGS model hold. In addition, df for the DPS[f] model is $(r-1)(r-2)/2$ and that for DGS model is $(r-1)$. The df for the S model can be obtained by adding the degrees of freedom for the DPS[f] model and the DGS model. Thus, we consider partitioning test statistics.

Theorem 2 confirms that the DPS[f] model is equivalent to the DPS model. Therefore, the maximum likelihood estimates (MLEs) under the DPS[f] model are given by

$$\begin{cases} \hat{m}_{ij} = \dfrac{n_k^U}{n_k^U + n_k^L}(n_{ij} + n_{ji}) & (i < j), \\ \hat{m}_{ij} = n_{ij} & (i = j), \\ \hat{m}_{ij} = \dfrac{n_k^L}{n_k^U + n_k^L}(n_{ij} + n_{ji}) & (i > j), \end{cases} \tag{7}$$

where $k = |j-i|$, $n_k^U = \sum\sum_{k=j-i} n_{ij}$, and $n_k^L = \sum\sum_{k=j-i} n_{ji}$ (Goodman [10]).

Next, we consider the MLEs under the DGS model using the Lagrange function. Since the kernel of the log likelihood is $\sum_{i=1}^{r}\sum_{j=1}^{r} n_{ij}\log\pi_{ij}$, Lagrange function L is written as

$$L = \sum_{i=1}^{r}\sum_{j=1}^{r} n_{ij}\log\pi_{ij} + \lambda\left(\sum_{i=1}^{r}\sum_{j=1}^{r}\pi_{ij} - 1\right) + \sum_{k=1}^{r-1}\lambda_k\left(\sum\sum_{k=j-i}(\pi_{ij} - \pi_{ji})\right).$$

Equating the derivation of L to 0 with respect to π_{ij}, λ, and λ_k gives

$$\begin{cases} \hat{m}_{ij} = \dfrac{(n_k^U + n_k^L)n_{ij}}{2n_k^U} & (i < j), \\ \hat{m}_{ij} = n_{ij} & (i = j), \\ \hat{m}_{ij} = \dfrac{(n_k^U + n_k^L)n_{ij}}{2n_k^L} & (i > j), \end{cases} \tag{8}$$

where $k = |j - i|$. It is important to note that the DPS and DGS models do not remain the same when the row and column categories are permuted. Therefore, these models should be used with data from an ordinal category.

We obtain the following equivalence from Equations (7) and (8):

$$G^2(S) = G^2(DPS[f]) + G^2(DGS),$$

because the MLEs under the S model are $\hat{m}_{ij} = (n_{ij} + n_{ji})/2$. Therefore, the DPS[$f$] model and the DGS model are separable and exhibit independence.

Let $W(M)$ denote the Wald statistic for model M. We obtain the following theorem and prove it in Appendix A.4.

Theorem 4. *$W(S)$ is equal to the sum of $W(DPS[f])$ and $W(DGS)$.*

4. Numerical Example

Table 1, which is taken from Smith et al. [22], describes the amount of influence religious leaders and medical leaders should have in government funding for decisions on stem cell research when surveying 871 people. The influence levels are divided into four categories: (1) Great influence, (2) Some influence, (3) A little influence, and (4) No influence.

Table 1. How much influence should religious leaders and medical leaders have in government funding for decisions on stem cell research? [22].

Religious Leaders	Medical Leaders: Great (1)	Fair (2)	Little (3)	None (4)	Total
Great (1)	36 (36.00) [a] (36.00) [b]	16 (11.96) [a] (60.19) [b]	7 (6.22) [a] (70.95) [b]	7 (7.00) [a] (67.00) [b]	66
Fair (2)	74 (78.04) [a] (42.67) [b]	96 (96.00) [a] (96.00) [b]	22 (26.05) [a] (82.76) [b]	4 (4.78) [a] (40.55) [b]	196
Little (3)	119 (119.78) [a] (62.59) [b]	174 (169.95) [a] (100.34) [b]	48 (48.00) [a] (48.00) [b]	4 (3.99) [a] (15.05) [b]	345
None (4)	127 (127.00) [a] (67.00) [b]	93 (92.22) [a] (48.91) [b]	26 (26.01) [a] (14.99) [b]	18 (18.00) [a] (18.00) [b]	264
Total	356	379	103	33	871

[a] MLEs under the DPS model; [b] MLEs under the DGS model.

The values of the likelihood ratio chi-square statistics G^2 and the corresponding p values for the models applied to these data are shown in Table 2. Table 2 indicates that the sum of the test statistics DPS (i.e., DPS[f]) model and DGS model is equal to that of the S model. The S model fits the data very poorly. We can infer that the marginal distribution for religious leaders is not equal to that for medical leaders. On the other hand, the DPS model fits the data very well. The likelihood-ratio test for the null hypothesis H_0: $d_1 = d_2 = d_3 = 1$ uses a test statistic which is the difference between G^2 for the S model and the DPS model. The resulting test statistic is $545.15 - 2.45 = 542.70$ with three degrees of freedom. This indicates strong evidence of at least one difference from 1. Additionally, the DGS model fits the data poorly. From Theorem 3, the reason of the poor fit of S model is caused by the poor fit of the DGS model rather than the DPS model.

Table 2. Likelihood ratio chi-square values G^2 for the models applied to Table 1.

Models	df	G^2	*p*-Value
S	6	545.15	<0.0001
DPS	3	2.45	0.4847
DGS	3	542.70	<0.0001

The values of MLEs of (d_1, d_2, d_3) in Equation (1) are $(0.15, 0.05, 0.06)$. It should be noted that (d_1, d_2, d_3) is equal to $(d_1^{\text{KL}}, d_2^{\text{KL}}, d_3^{\text{KL}})$ in the DPS_{KL} model. Let (i, j) denote the pair that the amount of influence religious leaders is ith level and that of medical leaders is jth level. When $k = j - i$ ($k = 1, 2, 3$), a pair (i, j) is $\hat{d}_k$ times as likely as a pair (j, i) on condition that a pair is (i, j) or (j, i). From $\hat{d}_k < 1$ ($k = 1, 2, 3$), the probability distribution for religious leaders is *stochastically higher* than the probability distribution of medical leaders. That is, the medical leaders rather than the religious leaders should have influence in government funding for decisions on stem cell research.

Moreover, from Theorem 2, we can obtain various interpretations. Since the DPS model holds, the DPS_{RKL}, DPS_{P}, and $\text{DPS}_{\text{PD}(\lambda)}$ models also hold. For example, we obtain

$$(\hat{d}_1^{\text{RKL}}, \hat{d}_2^{\text{RKL}}, \hat{d}_3^{\text{RKL}}) = (6.37, 19.22, 18.09),$$
$$(\hat{d}_1^{\text{P}}, \hat{d}_2^{\text{P}}, \hat{d}_3^{\text{P}}) = (-0.73, -0.90, -0.90),$$

and for $\lambda = 3$,

$$(\hat{d}_1^{\text{PD}(3)}, \hat{d}_2^{\text{PD}(3)}, \hat{d}_3^{\text{PD}(3)}) = (-0.65, -0.86, -0.85).$$

When $k = j - i$ ($k = 1, 2, 3$), we can infer that (i) the difference between the reciprocal of conditional probability that a pair is (i, j) and the reciprocal of conditional probability that a pair is (j, i) is $\hat{d}_k^{\text{RKL}}$ on condition that the pair is (i, j) or (j, i) from the DPS_{RKL} model, (ii) the difference between the conditional probability that a pair is (i, j) and the conditional probability that a pair is (j, i) is $\hat{d}_k^{\text{P}}$ under the same condition from the DPS_{P} model, and (iii) the difference between the conditional probability that a pair is (i, j) to the third power and the conditional probability that a pair is (j, i) to the third power is $\hat{d}_k^{\text{PD}(3)}$ under the same condition from the $\text{DPS}_{\text{PD}(3)}$ model.

5. Concluding Remarks

This paper proves that the DPS[f] model is equivalent to the DPS model proposed by Goodman [10]. This result provides various interpretations of the DPS model. The separation of the test statistic for the S model is considered. The DPS[f] and DGS models are separable and exhibit independence. Kateri and Papaioannou [2], Kateri and Agresti [3], Tahata [5] and Fujisawa and Tahata [14] considered models based on the f-divergence for the analysis of square contingency tables with ordinal categories. In the future, it should be studied whether the model based on the f-divergence is equivalent to the conventional model.

Author Contributions: Conceptualization, K.T.; methodology, K.T.; software, K.M.; validation, K.T. and K.M.; formal analysis, K.T.; investigation, K.M.; resources, Tahata, K.; data curation, K.M.; writing—original draft preparation, K.M.; writing—review and editing, K.T.; visualization, K.M.; supervision, K.T.; project administration, K.T.; funding acquisition, K.T. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by JSPS KAKENHI (Grant Number 20K03756).

Data Availability Statement: "The General Social Survey" at https://gss.norc.org/ (accessed on 1 June 2024).

Conflicts of Interest: The authors declare no conflicts of interest.

Appendix A

This section provides the proofs of theorems.

Appendix A.1

In a similar manner to Tahata [5], we prove Theorem 1. Let $I^C(\pi : \pi^S)$ denote the f-divergence between (π_{ij}) and (π_{ij}^S). That is

$$I^C(\pi : \pi^S) = \sum_{i=1}^{r}\sum_{j=1}^{r} \pi_{ij}^S f\left(\frac{\pi_{ij}}{\pi_{ij}^S}\right), \tag{A1}$$

where f satisfies the conditions described in Section 1. Now minimize (A1) under the conditions where the restraints

$$\pi_{ij} + \pi_{ji} = t_{ij} = t_{ji} \quad (i = 1, \dots, r; j = 1, \dots, r) \tag{A2}$$

and

$$\delta_{-k}^U = \sum\sum_{i-j=-k} \pi_{ij}, \quad \delta_k^L = \sum\sum_{i-j=k} \pi_{ij} \quad (k = 1, \dots, r-1) \tag{A3}$$

are given. The Lagrange function is written as

$$L = I^C(\pi : \pi^S) + \sum_{i=1}^{r}\sum_{j=1}^{r} \lambda_{ij}(\pi_{ij} + \pi_{ji} - t_{ij}) + \sum_{k=1}^{r-1}\left(\bar{\Delta}_{-k}\left(\sum\sum_{i-j=-k} \pi_{ij} - \delta_{-k}^U\right) + \bar{\Delta}_k\left(\sum\sum_{i-j=k} \pi_{ij} - \delta_k^L\right)\right).$$

By taking the partial derivative of L with respect to π_{ij} and setting it to zero, we obtain the following equation:

$$\begin{cases} f'\left(\frac{\pi_{ij}}{\pi_{ij}^S}\right) + \bar{\Delta}_{-k} + \lambda_{ij} + \lambda_{ji} = 0 & (i < j), \\ f'\left(\frac{\pi_{ij}}{\pi_{ij}^S}\right) + \lambda_{ij} + \lambda_{ji} = 0 & (i = j), \\ f'\left(\frac{\pi_{ij}}{\pi_{ij}^S}\right) + \bar{\Delta}_k + \lambda_{ij} + \lambda_{ji} = 0 & (i > j). \end{cases} \tag{A4}$$

Let f' denote F, and let π_{ij}^* denote the solution satisfying (A2), (A3), and (A4). Given that f is a strictly convex function, it follows that $F'(x) = f''(x) > 0$ for all x. Thus, F is strictly monotonic, ensuring the existence of F^{-1}. We represent ζ_{ij} as $-(\lambda_{ij} + \lambda_{ji})$ and Δ_l as $-\bar{\Delta}_l$. From Equation (A4), we obtain

$$\begin{cases} \pi_{ij}^* = \pi_{ij}^S F^{-1}(\Delta_{-k} + \zeta_{ij}) & (i < j), \\ \pi_{ij}^* = \pi_{ij}^S F^{-1}(\zeta_{ij}) & (i = j), \\ \pi_{ij}^* = \pi_{ij}^S F^{-1}(\Delta_k + \zeta_{ij}) & (i > j), \end{cases}$$

where $\zeta_{ij} = \zeta_{ji}$ and $\Delta_k + \Delta_{-k} = 0$. The minimum value of $I^C(\pi : \pi^S)$ is obtained for π_{ij}^*, where ζ_{ij} and Δ_l are selected to ensure that π_{ij}^* complies with the constraints (A2) and (A3). Thus, the DPS[f] model represents the optimal approximation to the S model in terms of f-divergence under these specified conditions.

Appendix A.2

Let function G be defined as

$$G(x) = F\left(\frac{2x}{1+x}\right) - F\left(\frac{2}{1+x}\right) \quad (x > 0),$$

where $F = f'$. Then, the derivative of G is

$$G'(x) = \frac{2}{(1+x)^2}\left(F'\left(\frac{2x}{1+x}\right) + F'\left(\frac{2}{1+x}\right)\right).$$

Since the function f is twice-differential and strictly convex $G'(x) > 0$ for $x > 0$, hence G is a strictly increasing function, and G^{-1} exists.

If the DPS model holds, $\pi_{ij}/\pi_{ji} = d_k$ holds for $i < j$ from Equation (1), where $k = j - i$. Then we can see that for $i < j$,

$$\begin{aligned} G(d_k) &= F\left(\frac{2d_k}{1+d_k}\right) - F\left(\frac{2}{1+d_k}\right), \\ &= F\left(2\pi^c_{ij}\right) - F\left(2\pi^c_{ji}\right). \end{aligned}$$

This is equivalent to Equation (4). Namely, the DPS[f] model holds.

On the other hand, if the DPS[f] model holds, Equation (4) holds. We can see that for $i < j$,

$$G\left(\frac{\pi_{ij}}{\pi_{ji}}\right) = a_k.$$

Since G^{-1} exists, we obtain

$$\frac{\pi_{ij}}{\pi_{ji}} = G^{-1}(a_k).$$

Namely, the DPS model holds. The proof is complete.

Appendix A.3

It is obvious that if the S model holds, the DPS[f] model and the DGS model simultaneously hold. Assuming that both the DPS[f] and the DGS models hold, we show that the S model holds. From Theorem 2, the DPS[f] model is equivalent to $\pi_{ij}/\pi_{ji} = d_k$ for $i < j$ with $k = j - i$. Since the DGS model holds, we obtain

$$\sum_{j-i=k}\sum (d_k - 1)\pi_{ji} = 0 \quad (k = 1, \ldots, r-1).$$

Since $\pi_{ji} > 0$, we get $d_k = 1$ ($k = 1, \ldots, r-1$). Namely, the S model holds.

Appendix A.4

Theorem 2 shows that the DPS[f] model is equivalent to the DPS model. Let

$$\begin{aligned} \boldsymbol{\pi} &= (\pi_{11}, \ldots, \pi_{1r}, \pi_{21}, \ldots, \pi_{2r}, \ldots, \pi_{r1}, \ldots, \pi_{rr})^t, \\ \boldsymbol{\beta} &= (\rho_1, \ldots, \rho_{r-1}, \boldsymbol{\varepsilon})^t, \end{aligned}$$

where $\boldsymbol{\varepsilon} = (\varepsilon_{11}, \ldots, \varepsilon_{1r}, \varepsilon_{22}, \ldots, \varepsilon_{2r}, \ldots, \varepsilon_{rr})$. Then, from Equation (1), the DPS model is expressed as

$$\log \boldsymbol{\pi} = \boldsymbol{X}\boldsymbol{\beta} = (\boldsymbol{x}_1, \ldots, \boldsymbol{x}_{r-1}, \boldsymbol{x}_{11}, \ldots, \boldsymbol{x}_{1r}, \boldsymbol{x}_{22}, \ldots, \boldsymbol{x}_{2r}, \ldots, \boldsymbol{x}_{rr})\boldsymbol{\beta},$$

where $\boldsymbol{x}_l = (\boldsymbol{w}_{l+1}, \ldots, \boldsymbol{w}_r, 0, \ldots, 0)^t$ is a $r^2 \times 1$ vector ($l = 1, \ldots, r-1$). Here, $\boldsymbol{w}_h$ ($1 \times r$ vector) is 1 for the hth element and 0 otherwise. For example, when $r = 4$,

$$\boldsymbol{x}_1 = (\boldsymbol{w}_2, \boldsymbol{w}_3, \boldsymbol{w}_4, 0, \ldots, 0)^t = (0,1,0,0,0,0,1,0,0,0,0,1,0,0,0,0)^t.$$

Additionally, $\boldsymbol{x}_{ij}$ ($i \leq j$) is the $r^2 \times 1$ vector shouldering ε_{ij}. Note that the $r^2 \times K$ matrix $\boldsymbol{X}$ is a full column rank where $K = (r-1) + r(r+1)/2$.

We define the linear space spanned by the columns of the matrix $\boldsymbol{X}$ as $S(\boldsymbol{X})$, which has dimension K. This space, $S(\boldsymbol{X})$, is a subspace of $\mathbb{R}^{r^2}$. Consider an $r^2 \times d_1$ matrix $\boldsymbol{U}$ with full column rank, such that the linear space $S(\boldsymbol{U})$, spanned by the columns of $\boldsymbol{U}$, serves as the orthogonal complement of $S(\boldsymbol{X})$. Note that d_1 is calculated as $d_1 = r^2 - ((r-1) + r(r+1)/2) = (r-1)(r-2)/2$. Given that $\boldsymbol{U}^t\boldsymbol{X} = \boldsymbol{O}_{d_1,K}$, where $\boldsymbol{O}_{d_1,K}$ denotes the $d_1 \times K$ zero matrix, the DPS model can be expressed as $\boldsymbol{h}_1(\boldsymbol{\pi}) = \boldsymbol{U}^t \log \boldsymbol{\pi} = \boldsymbol{0}_{d_1}$, with $\boldsymbol{0}_s$ representing the $s \times 1$ zero vector.

Additionally, the DGS model can be expressed as $\boldsymbol{h}_2(\boldsymbol{\pi}) = \boldsymbol{M}\boldsymbol{\pi} = \boldsymbol{0}_{d_2}$ where

$$\boldsymbol{M} = (\boldsymbol{g}_1, \ldots, \boldsymbol{g}_{r-1})^t,$$

and $d_2 = r - 1$. Here, $\boldsymbol{g}_l = 2\boldsymbol{x}_l - \sum\sum_{j-i=l} \boldsymbol{x}_{ij}$. Note that $\boldsymbol{M}^t$ belongs to the space $S(\boldsymbol{X})$. That is, $S(\boldsymbol{M^t}) \subset S(\boldsymbol{X})$.

Let $\boldsymbol{p}$ denote $\boldsymbol{\pi}$ with π_{ij} replaced by p_{ij}, where $p_{ij} = n_{ij}/n$ with $n = \sum\sum n_{ij}$. From Theorem 3, the S model is equivalent to $\boldsymbol{h}_3(\boldsymbol{\pi}) = \boldsymbol{0}_{d_3}$, where $\boldsymbol{h}_3 = (\boldsymbol{h}_1^t, \boldsymbol{h}_2^t)^t$ and $d_3 = d_1 + d_2 = r(r-1)/2$. In an analogous manner to Tahata [5], we obtain that $\sqrt{n}(\boldsymbol{h}_3(\boldsymbol{p}) - \boldsymbol{h}_3(\boldsymbol{\pi}))$ has an asymptotically normal distribution with mean $\boldsymbol{0}_{d_3}$ and covariance matrix

$$\boldsymbol{H}_3(\boldsymbol{\pi})\boldsymbol{\Sigma}(\boldsymbol{\pi})\boldsymbol{H}_3^t(\boldsymbol{\pi}) = \begin{bmatrix} \boldsymbol{H}_1(\boldsymbol{\pi})\boldsymbol{\Sigma}(\boldsymbol{\pi})\boldsymbol{H}_1^t(\boldsymbol{\pi}) & \boldsymbol{O}_{d_1,d_2} \\ \boldsymbol{O}_{d_2,d_1} & \boldsymbol{H}_2(\boldsymbol{\pi})\boldsymbol{\Sigma}(\boldsymbol{\pi})\boldsymbol{H}_2^t(\boldsymbol{\pi}) \end{bmatrix},$$

where $\boldsymbol{H}_s(\boldsymbol{\pi}) = \partial \boldsymbol{h}_s(\boldsymbol{\pi})/\boldsymbol{\pi}^t$ and $\boldsymbol{\Sigma}(\boldsymbol{\pi}) = diag(\boldsymbol{\pi}) - \boldsymbol{\pi}\boldsymbol{\pi}^t$. Here, $diag(\boldsymbol{\pi})$ denotes a diagonal matrix with the ith component of $\boldsymbol{\pi}$ as the ith diagonal component. Therefore, $W_3 = W_1 + W_2$ holds, where

$$W_s = n\boldsymbol{h}_s^t(\boldsymbol{p})(\boldsymbol{H}_s(\boldsymbol{p})\boldsymbol{\Sigma}(\boldsymbol{p})\boldsymbol{H}_s^t(\boldsymbol{p}))^{-1}\boldsymbol{h}_s(\boldsymbol{p}).$$

The Wald statistic for the DPS[f] model (i.e., $W(DPS[f])$) is W_1, that for the DGS model (i.e., $W(DGS)$) is W_2, and that for the S model (i.e., $W(S)$) is W_3. The proof is complete.

References

1. Bowker, A.H. A test for symmetry in contingency tables. *J. Am. Stat. Assoc.* **1948**, *43*, 572–574. [CrossRef]
2. Kateri, M.; Papaioannou, T. Asymmetry models for contingency tables. *J. Am. Stat. Assoc.* **1997**, *92*, 1124–1131. [CrossRef]
3. Kateri, M.; Agresti, A. A class of ordinal quasi-symmetry models for square contingency tables. *Stat. Probab. Lett.* **2007**, *77*, 598–603. [CrossRef]
4. Tahata, K.; Tomizawa, S. Generalized linear asymmetry model and decomposition of symmetry for multiway contingency tables. *J. Biom. Biostat.* **2011**, *2*, 1–6. [CrossRef]
5. Tahata, K. Separation of symmetry for square tables with ordinal categorical data. *Jpn. J. Stat. Data Sci.* **2020**, *3*, 469–484. [CrossRef]
6. Tahata, K. Advances in Quasi-Symmetry for Square Contingency Tables. *Symmetry* **2022**, *14*, 1051. [CrossRef]
7. Beh, E.J.; Lombardo, R. Visualising Departures from Symmetry and Bowker's X2 Statistic. *Symmetry* **2022**, *14*, 1103. [CrossRef]
8. Altun, G.; Saraçbaşı, T. Determination of model fitting with power-divergence-type measure of departure from symmetry for sparse and non-sparse square contingency tables. *Commun. Stat.-Simul. Comput.* **2022**, *51*, 4087–4111. [CrossRef]
9. Ando, S. Generalized Sum-Asymmetry Model and Orthogonality of Test Statistic for Square Contingency Tables. *Austrian J. Stat.* **2024**, *53*, 99–108. [CrossRef]
10. Goodman, L.A. Multiplicative Models for Square Contingency Tables with Ordered Categories. *Biometrika* **1979**, *66*, 413–418. [CrossRef]
11. McCullagh, P. A class of parametric models for the analysis of square contingency tables with ordered categories. *Biometrika* **1978**, *65*, 413–418. [CrossRef]

12. Ireland, C.T.; Ku, H.H.; Kullback, S. Symmetry and Marginal Homogeneity of an $r \times r$ Contingency Table. *J. Am. Stat. Assoc.* **1969**, *64*, 1323–1341. [CrossRef]
13. Read, C.B.; Cressie, N. *Goodness-of-Fit Statistics for Discrete Multivariate Data*; Springer: New York, NY, USA, 1988.
14. Fujisawa, K.; Tahata, K. Asymmetry model based on f-divergence and orthogonal decomposition of symmetry for square contingency tables with ordinal categories. *SUT J. Math.* **2020**, *56*, 39–53. [CrossRef]
15. Read, C.B. Partitioning chi-squape in contingency tables: A teaching approach. *Commun. Stat.-Theory Methods* **1977**, *6*, 553–562. [CrossRef]
16. Tahata, K.; Naganawa, M.; Tomizawa, S. Extended linear asymmetry model and separation of symmetry for square contingency tables. *J. Jpn. Stat. Soc.* **2016**, *46*, 189–202. [CrossRef]
17. Tahata, K.; Auchi, R.; Ando, S.; Tomizawa, S. Separation of the refined estimator of the measure for symmetry in square contingency tables. *Commun. Stat.-Simul. Comput.* **2023**, 1–17. [CrossRef]
18. Aitchison, J. Large-sample restricted parametric tests. *J. R. Stat. Soc. Ser. B-Stat. Methodol.* **1962**, *24*, 234–250. [CrossRef]
19. Darroch, J.N.; Silvey, S.D. On testing more than one hypothesis. *Ann. Math. Stat.* **1963**, *34*, 555–567. [CrossRef]
20. Lang, J.B.; Agresti, A. Simultaneously modeling joint and marginal distributions of multivariate categorical responses. *J. Am. Stat. Assoc.* **1994**, *89*, 625–632. [CrossRef]
21. Lang, J.B. On the partitioning of goodness-of-fit statistics for multivariate categorical response models. *J. Am. Stat. Assoc.* **1996**, *91*, 1017–1023. [CrossRef]
22. Smith, T.W.; Marsden, P.; Hout, M.; Kim, J. *General Social Surveys, 1972–2014 [Machine-Readable Data File]*; NORC at the University of Chicago: Chicago, IL, USA, 2006.