SUT Journal of Mathematics
Vol. 56, No. 1 (2020), 39–53

Asymmetry model based on f-divergence and orthogonal decomposition of symmetry for square contingency tables with ordinal categories

Kengo Fujisawa and Kouji Tahata

(Received February 25, 2020)

Abstract. Many statisticians have considered various symmetry and asymmetry models to analyze square contingency tables with ordinal categories. We propose a generalized model, which indicates the asymmetric structure for cell probabilities. This model is the closest to the symmetry model in terms of the f-divergence under certain conditions. Then we decompose the symmetry model using the proposed model, and partition the likelihood-ratio chi-square test statistics for the symmetry model. The decomposition is useful to deduce the reason for a poor fit of the symmetry model.

AMS 2010 *Mathematics Subject Classification.* 62H17.

Key words and phrases. Square contingency table, f-divergence, ordinal categories.

§1. Introduction

In a square contingency table with ordinal row and column classifications, determining the structure symmetry is important since the independence between the row and column classifications is unlikely to hold. Many statisticians have proposed various symmetry and asymmetry models such as those by Bowker (1948), Stuart (1955), Caussinus (1965), Bhapkar (1966), and McCullagh (1978). Moreover, Kateri and Papaioannou (1997), Kateri and Agresti (2007), and Tahata (2019) proposed asymmetry models based on the f-divergence. In this paper, we propose an asymmetry model based on the f-divergence, which generalizes various symmetry and asymmetry models, and show that the proposed model is the closest to the symmetry model in terms of the f-divergence under certain conditions.

Caussinus (1965) provided a theorem that the symmetry model holds if and only if both the symmetry of the odds ratios and the homogeneity of marginal distribution hold. A theorem may be useful to determine the reason for a poor fit of the symmetry model. Hence, we also consider the decomposition of symmetry using our proposed model.

Aitchison (1962), Darroch and Silvey (1963), Lang and Agresti (1994), and Tomizawa and Tahata (2007) argued for orthogonality of the model. It means that a test statistic for the goodness-of-fit of model M1 is asymptotically equivalent to the sum of the test statistics for model M2 and model M3 when model M1 can be decomposed into model M2 and model M3. When it holds, an incompatible situation, where both model M2 and model M3 are accepted but model M1 is rejected, would not arise. In this paper, we give the orthogonal decomposition where the test statistic for the symmetry model is equal to the sum of those for the decomposed models.

The present paper is organized as follows. Section 2 proposes our new model. Section 3 gives the decomposition of the symmetry model. Section 4 partitions the test statistic. Section 5 gives numerical examples. Section 6 concludes this paper.

§2. Models

Consider an $r \times r$ square contingency table with the same row and column classifications. Let π_{ij} denote the probability that an observation will fall in the ith row and the jth column of the table ($i = 1, \ldots, r; j = 1, \ldots, r$), and let π_{i+} and π_{+i} denote the row and column marginal probabilities ($i = 1, \ldots, r$), respectively. Bowker (1948) proposed the symmetry (S) model, which is defined as

$$\pi_{ij} = \pi_{ji} \quad (i = 1, \ldots, r; j = 1, \ldots, r).$$

This indicates that the probability of an observation falling in the (i, j) cell, $i \neq j$, is equal to the probability of the observation falling in the symmetric (j, i) cell. As a model that has a weaker restriction, Caussinus (1965) considered the quasi-symmetry (QS) model, which is defined as

$$\pi_{ij} = \mu \alpha_i \beta_j \psi_{ij} \quad (i = 1, \ldots, r;\ j = 1, \ldots, r),$$

where $\psi_{ij} = \psi_{ji}$. The QS model indicates the symmetry with respect to the odds ratios. Note that the QS model with $\{\alpha_i = \beta_i\}$ is the S model.

When the S model fits a given dataset poorly, applying an asymmetry model may be more appropriate. For a square contingency table with ordinal categories, McCullagh (1978) proposed the conditional symmetry (CS) model, which is defined as

$$\pi_{ij} = \gamma \pi_{ji} \quad (i < j).$$

The CS model indicates that the probability of an observation falling in the (i,j) cell, $i<j$, is γ times higher than that in the symmetric (j,i) cell. The CS model with $\gamma = 1$ is the S model.

Agresti (1983) proposed the linear diagonals-parameter symmetry (LDPS) model, which is defined as

$$\pi_{ij} = \theta^{j-i}\pi_{ji} \quad (i<j).$$

The LDPS model indicates that the ratios of symmetric cells increase or decrease exponentially as the difference $j-i$. A special case of this model with $\theta = 1$ is the S model.

Tahata, Naganawa, and Tomizawa (2016) proposed the extended kth linear asymmetry (ELS_k) model, which is defined for a fixed k $(k = 1,\dots,r-1)$ as

$$\pi_{ij} = \left(\gamma \prod_{l=1}^{k} \theta_l^{j^l - i^l}\right)\pi_{ji} \quad (i<j).$$

The ELS_k model indicates that the log-ratios of symmetric cells are expressed as the polynomial. The ELS_k model includes various models such as those proposed by Bowker (1948), McCullagh (1978), Agresti (1983), and Tomizawa (1987).

Let $p = (p_{ij})$ and $q = (q_{ij})$ be two discrete finite bivariate probability distributions. The f-divergence between p and q is defined as

$$I(p||q) = \sum_i \sum_j q_{ij} f\left(\frac{p_{ij}}{q_{ij}}\right),$$

where f is a convex function on $(0,\infty)$ with $f(1) = 0$, $f(0) = \lim_{t\to 0} f(t)$, $0\cdot f(0/0) = 0$, and $0 \cdot f(a/0) = a \lim_{t\to\infty}(f(t)/t)$. For details, see Csiszár and Shields (2004). Many divergences, such as the Kullback-Leibler divergence, Pearson χ^2 divergence, and power divergence (Read and Cressie, 1988) are special cases of the f-divergence.

Let f be a twice-differentiable and strictly convex function, and let $F(t) = df(t)/dt$ for all t. Kateri and Papaioannou (1997) introduced the quasi-symmetry (QS[f]) model based on the f-divergence defined as

$$\pi_{ij} = \pi_{ij}^S F^{-1}(\alpha_i + \gamma_{ij}) \quad (i = 1,\dots,r;\ j = 1,\dots,r),$$

where $\gamma_{ij} = \gamma_{ji}$ and $\pi_{ij}^S = (\pi_{ij}+\pi_{ji})/2$. It is proved that the QS[f] model is the closest model to the S model in terms of the f-divergence under the conditions where the row or column marginals π_{i+} (or π_{+i}) for $i = 1,\dots,r$ as well as the sums $\pi_{ij} + \pi_{ji}$ for $i = 1,\dots,r; j = 1,\dots,r$ are given. If $f(t) = t\log t$, $t>0$, then the f-divergence is reduced to the Kullback-Leibler divergence, and the

QS[f] model is equivalent to the QS model. Under these conditions, the QS model is the closest model to the S model in terms of the Kullback-Leibler divergence.

Let $\{u_i\}$ denote the known scores assigned to the row and column categories where $u_1 \leq u_2 \leq \cdots \leq u_r$ and $u_1 < u_r$. Kateri and Agresti (2007) proposed the ordinal quasi-symmetry (OQS[f]) model, which is based on the f-divergence and is defined as

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

where $\gamma_{ij} = \gamma_{ji}$ and $\pi_{ij}^S = (\pi_{ij} + \pi_{ji})/2$. The OQS[$f$] model is the closest model to the S model in terms of the f-divergence under the conditions that the sums $\pi_{ij} + \pi_{ji}$ for $i = 1, \ldots, r; j = 1, \ldots, r$ and marginal mean $\sum_i u_i \pi_{i+}$ (or $\sum_i u_i \pi_{+i}$) are given. The OQS[f] model is a special case of the QS[f] model.

We propose an asymmetry model based on the f-divergence. The conditional symmetry (CS[f]) model based on the f-divergence is defined as

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

where $\gamma_{ij} = \alpha_0 + \gamma_{ji}$ $(i < j)$ and $\pi_{ij}^S = (\pi_{ij} + \pi_{ji})/2$. The CS model is a special case of the CS[f] model. In an analogous manner to Kateri and Papaioannou (1997), we can obtain the following theorem:

Theorem 2.1. *The CS[f] model is the closest to the S model in terms of the f-divergence under the conditions where the sums $\{\pi_{ij} + \pi_{ji}\}$ and $\sum\sum_{i<j} \pi_{ij}$ (or $\sum\sum_{i<j} \pi_{ji}$) are given.*

Proof. We assume that the total upper-diagonal cell probabilities are given. $I(\pi || \pi^S)$ is minimized under the constraints of

$$\sum\sum_{i<j} \pi_{ij} = \delta,$$

and

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

where δ and $\{t_{ij}\}$ are constants. Note that $\{\pi_{ij}^S\}$ satisfies the structure of symmetry. Namely, $\pi_{ij}^S = \pi_{ji}^S$. This is a constraint minimization problem, which can be solved by the method of Lagrange multipliers. The Lagrange function is

$$L = I(\pi || \pi^S) + \psi \left(\sum\sum_{i<j} \pi_{ij} - \delta \right) + \sum_{i=1}^{r} \sum_{j=1}^{r} \phi_{ij} (\pi_{ij} + \pi_{ji} - t_{ij}),$$

where ψ and $\{\phi_{ij}\}$ are the Lagrange multipliers. Equating the partial derivative of L with respect to π_{ij} to 0, we obtain

$$\begin{cases} F\left(\dfrac{\pi_{ij}}{\pi_{ij}^S}\right) + \phi_{ij} + \phi_{ji} + \psi = 0 \quad (i < j), \\ \\ F\left(\dfrac{\pi_{ij}}{\pi_{ij}^S}\right) + \phi_{ij} + \phi_{ji} = 0 \quad (i \geq j). \end{cases}$$

The existence of F^{-1} is ensured because f is strictly convex. With $\alpha_0 = -\psi$ and $\gamma_{ij} = -(\phi_{ij} + \phi_{ji})$, we obtain the CS[$f$] model. The proof is complete. □

Let $\{u_i\}$ denote the known scores $u_1 < u_2 < \cdots < u_r$ and also let g_{ik} denote the function of $(\alpha_0, \ldots, \alpha_k)$ for a fixed k $(k = 0, \ldots, r-1)$, where

$$g_{ik}(\boldsymbol{\alpha}) = \begin{cases} 0 & (k = 0), \\ \displaystyle\sum_{h=1}^{k} u_i^h \alpha_h & (k = 1, \ldots, r-1). \end{cases}$$

As an extension of the CS[f] model, we also propose a new model defined for a fixed k $(k = 0, \ldots, r-1)$ as

$$\pi_{ij} = \pi_{ij}^S F^{-1}\left(g_{ik}(\boldsymbol{\alpha}) + \gamma_{ij}\right) \quad (i = 1, \ldots, r;\ j = 1, \ldots, r),$$

where $\gamma_{ij} = \alpha_0 + \gamma_{ji}$ $(i < j)$ and $\pi_{ij}^S = (\pi_{ij} + \pi_{ji})/2$. We refer to this model as the extended kth asymmetry model based on the f-divergence denoted by EA$_k$[f]. We note that the EA$_0$[f] model is the CS[f] model. As a special case of this model, the EA$_k$[f] model with $\alpha_0 = 0$ is the kth asymmetry model based on the f-divergence proposed by Tahata (2019). Note that (i) the EA$_k$[f] model is the S model when $\alpha_0 = \alpha_1 = \cdots = \alpha_k = 0$ for any f, (ii) the EA$_k$[f] model is the QS[f] model when $k = r-1$ and $\alpha_0 = 0$, and (iii) the EA$_k$[f] model is the OQS[f] model when $k = 1$ and $\alpha_0 = 0$.

The EA$_k$[f] model can be expressed as

$$F(2\pi_{ij}^c) = g_{ik}(\boldsymbol{\alpha}) + \gamma_{ij} \quad (i = 1, \ldots, r;\ j = 1, \ldots, r),$$

where $\gamma_{ij} = \alpha_0 + \gamma_{ji}$ $(i < j)$ and $\pi_{ij}^c = \pi_{ij}/(\pi_{ij} + \pi_{ji})$. Note that π_{ij}^c is the conditional probability of an observation falling in the (i, j) cell under the condition that the observation falls in the (i, j) cell or (j, i) cell.

If $f(t) = t \log t$, $t > 0$, then the f-divergence is reduced to the Kullback-Leibler divergence, $F^{-1}(t) = \exp(t - 1)$, and the EA$_k$[f] model becomes

$$\pi_{ij} = \pi_{ij}^S \exp\left(g_{ik}(\boldsymbol{\alpha}) + \gamma_{ij} - 1\right) \quad (i = 1, \ldots, r;\ j = 1, \ldots, r), \tag{2.1}$$

where $\gamma_{ij} = \alpha_0 + \gamma_{ji}$ $(i < j)$ and $\pi_{ij}^S = (\pi_{ij} + \pi_{ji})/2$. From equation (2.1), we obtain that for $i < j$,

$$\frac{\pi_{ij}}{\pi_{ji}} = \begin{cases} \beta_0 & (k = 0), \\ \beta_0 \displaystyle\prod_{h=1}^{k} \beta_h^{u_i^h - u_j^h} & (k = 1, \ldots, r-1), \end{cases} \tag{2.2}$$

where $\beta_l = \exp(\alpha_l)$ $(l = 0, 1, \ldots, k)$. Equation (2.2) with $\{u_i = i\}$ is identical to the ELS_k model. Therefore, the $\mathrm{EA}_k[f]$ model generalizes the aforementioned models.

If $f(t) = (1-t)^2$, then $F^{-1}(t) = (t/2)+1$, and the $\mathrm{EA}_k[f]$ model is reduced to

$$\pi_{ij} = \pi_{ij}^S \left(\frac{g_{ik}(\boldsymbol{\alpha}) + \gamma_{ij}}{2} + 1 \right) \quad (i = 1, \ldots, r;\ j = 1, \ldots, r), \tag{2.3}$$

where $\gamma_{ij} = \alpha_0 + \gamma_{ji}$ $(i < j)$ and $\pi_{ij}^S = (\pi_{ij} + \pi_{ji})/2$. This model can be expressed as, for $i < j$,

$$\pi_{ij}^c - \pi_{ji}^c = \begin{cases} \beta_0 & (k = 0), \\ \beta_0 + \displaystyle\sum_{h=1}^{k} (u_i^h - u_j^h)\beta_h & (k = 1, \ldots, r-1), \end{cases}$$

where $\beta_l = \alpha_l/4$ $(l = 0, 1, \ldots, k)$. This model is based on the Pearson χ^2 divergence. We refer to this model as the Pearsonian extended kth asymmetry (PEA_k) model.

Moreover, consider $f(t) = (\lambda(\lambda+1))^{-1}(t^{\lambda+1} - t)$, $t > 0$, where λ is a real-valued parameter. It is defined by the continuous limits as $\lambda \to 0$ and $\lambda \to -1$ for $\lambda = 0$ and $\lambda = -1$. In this case, the f-divergence is reduced to the power divergence, $F^{-1}(t) = (\lambda t + (\lambda+1)^{-1})^{1/\lambda}$ for $\lambda \neq 0, -1$, and the $\mathrm{EA}_k[f]$ model becomes

$$\pi_{ij} = \pi_{ij}^S (\lambda(g_{ik}(\boldsymbol{\alpha}) + \gamma_{ij}) + (\lambda+1)^{-1})^{\frac{1}{\lambda}} \quad (i = 1, \ldots, r;\ j = 1, \ldots, r),$$

where $\gamma_{ij} = \alpha_0 + \gamma_{ji}$ $(i < j)$ and $\pi_{ij}^S = (\pi_{ij} + \pi_{ji})/2$. This model can be expressed as

$$\pi_{ij} = \begin{cases} \pi_{ij}^S (1 + \lambda\psi_{ij})^{\frac{1}{\lambda}} & (k = 0), \\ \pi_{ij}^S \left(1 + \lambda \left(\displaystyle\sum_{h=1}^{k} u_i^h \alpha_h + \psi_{ij} \right)\right)^{\frac{1}{\lambda}} & (k = 1, \ldots, r-1), \end{cases} \tag{2.4}$$

for $i = 1, \ldots, r;\ j = 1, \ldots, r$, where $\psi_{ij} = \gamma_{ij} - (\lambda+1)^{-1}$ and $\psi_{ij} = \alpha_0 + \psi_{ji}$ $(i < j)$. Equation (2.4) with $\lambda \to 0$ can be expressed as equation (2.1), and equation (2.4) with $\lambda = 1$ is equivalent to equation (2.3) by replacing $\alpha_h = \alpha'_h/2$ and $\psi_{ij} = \psi'_{ij}/2$.

Similarly, we obtain the following theorem with respect to the $\mathrm{EA}_k[f]$ model:

Theorem 2.2. *For any k ($k = 1, \ldots, r-1$), the $\mathrm{EA}_k[f]$ model is the closest to the S model in terms of the f-divergence under the conditions that the sums $\{\pi_{ij} + \pi_{ji}\}$, the hth moment $\sum_i u_i^h \pi_{i+}$ (or $\sum_i u_i^h \pi_{+i}$) ($h = 1, \ldots, k$), and $\sum\sum_{i<j} \pi_{ij}$ (or $\sum\sum_{i<j} \pi_{ji}$) are given.*

From Theorem 2.2, the ELS_k model is the closest model to the S model in terms of the Kullback-Leibler divergence under these conditions.

§3. Decomposition of symmetry

Let X_1 and X_2 denote the row and column variables, respectively, and let $\{u_s\}$ ($s = 1, \ldots, r$) be a set of known scores. We consider a model defined for a fixed k ($k = 1, \ldots, r-1$) as

$$E(X_1^h) = E(X_2^h) \quad (h = 1, \ldots, k),$$

where $E(X_1^h) = \sum_i u_i^h \pi_{i+}$ and $E(X_2^h) = \sum_j u_j^h \pi_{+j}$. We refer to this model as the marginal kth moment equality (ME_k) model.

The global symmetry (GS) model is defined as

$$\delta_U = \delta_L,$$

where $\delta_U = \sum\sum_{i<j} \pi_{ij}$ and $\delta_L = \sum\sum_{i<j} \pi_{ji}$ (see Read, 1977). Then, we obtain the following theorem.

Theorem 3.1. *For any k ($k = 1, \ldots, r-1$), the S model holds if and only if the $\mathrm{EA}_k[f]$, GS, and ME_k models simultaneously hold.*

Proof. If the S model holds, then for any k, the $\mathrm{EA}_k[f]$, GS, and ME_k models simultaneously hold. Assuming that all the $\mathrm{EA}_k[f]$, GS, and ME_k models hold for some k, then we prove that the S model holds. Let $\{\tilde{\pi}_{ij}\}$ denote the cell probabilities satisfying all the $\mathrm{EA}_k[f]$, GS, and ME_k models hold. From the $\mathrm{EA}_k[f]$ model, we obtain

$$F\left(\frac{\tilde{\pi}_{ij}}{\tilde{\pi}_{ij}^S}\right) = \sum_{h=1}^{k} u_i^h \alpha_h + \gamma_{ij} \quad (i = 1, \ldots, r;\ j = 1, \ldots, r),$$

where $\gamma_{ij} = \alpha_0 + \gamma_{ji}$ ($i < j$) and $\tilde{\pi}_{ij}^S = (\tilde{\pi}_{ij} + \tilde{\pi}_{ji})/2$. From this equation, we obtain

$$F\left(\frac{\tilde{\pi}_{ij}}{\tilde{\pi}_{ij}^S}\right) - F\left(\frac{\tilde{\pi}_{ji}}{\tilde{\pi}_{ji}^S}\right) = \alpha_0 + \sum_{h=1}^{k}(u_i^h - u_j^h)\alpha_h \quad (i < j).$$

The ME_k model can be expressed as

$$\mu_1^h = \mu_2^h,$$

where $\mu_1^h = \sum\sum_{i<j}(u_i^h - u_j^h)\pi_{ij}$ and $\mu_2^h = \sum\sum_{i<j}(u_i^h - u_j^h)\pi_{ji}$. Assuming that the GS and ME_k models hold, we can see that

$$\sum\sum_{i<j} (\tilde{\pi}_{ij} - \tilde{\pi}_{ji}) \left(F\left(\frac{\tilde{\pi}_{ij}}{\tilde{\pi}_{ij}^S}\right) - F\left(\frac{\tilde{\pi}_{ji}}{\tilde{\pi}_{ji}^S}\right)\right) = 0.$$

Therefore, the S model holds because the monotonicity of F is ensured. □

Note that this theorem generalizes the results of Caussinus (1965), Kateri and Papaioannou (1997), and Tahata *et al.* (2016).

When $k = 0$, we can obtain the following corollary.

Corollary 3.1. *The S model holds if and only if both the* $EA_0[f]$ *and GS models hold.*

§4. Partition of test statistics

For the $r \times r$ contingency tables, let n_{ij} denote the observed frequency in the (i, j)th cell of the table and m_{ij} denote the corresponding expected frequency with $n = \sum\sum n_{ij}$ $(i = 1, \dots, r; j = 1, \dots, r)$. Assume that $\{n_{ij}\}$ has a multinomial distribution. $\{\hat{m}_{ij}\}$ denotes the maximum likelihood estimate (MLE) of $\{m_{ij}\}$ under a model. The likelihood ratio chi-squared statistic for the goodness-of-fit of the model M is defined as

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

The numbers of degrees of freedom (df) for testing the goodness-of-fit under the $\mathrm{EA}_k[f]$, GS, and ME_k models are $r(r-1)/2 - k - 1$, 1, and k, respectively. The number of df for the S model is equal to the sum of the numbers of df for the $\mathrm{EA}_k[f]$, GS, and ME_k models.

Considering the model where both the GS and ME_k models hold, which is denoted by GM_k, we can obtain the following theorem.

Theorem 4.1. *Under the S model, for any* k $(k = 1, \dots, r-1)$*, the test statistic* $G^2(S)$ *is asymptotically equivalent to the sum of* $G^2(EA_k[f])$ *and* $G^2(GM_k)$*.*

Proof. For a fixed k $(k = 1, \dots, r-1)$, the $\text{EA}_k[f]$ model may be expressed as

$$F(2\pi_{ij}^c) = \sum_{h=1}^{k}(u_i^h - u_j^h)\phi_h + \gamma_{ij} \quad (i = 1, \dots, r;\ j = 1, \dots, r),$$

where $\gamma_{ij} = \alpha_0 + \gamma_{ji}$ $(i < j)$ and $\phi_h = \alpha_h/2$ $(h = 1, \dots, k)$. Let

$$\boldsymbol{\pi} = (\pi_{11}, \dots, \pi_{1r}, \pi_{21}, \dots, \pi_{2r}, \dots, \pi_{rr})',$$

$$\boldsymbol{F}(2\boldsymbol{\pi}^c) = (F(2\pi_{11}^c), \dots, F(2\pi_{1r}^c), F(2\pi_{21}^c), \dots, F(2\pi_{2r}^c), \dots, F(2\pi_{rr}^c))',$$

$$\boldsymbol{\beta} = (\alpha_0, \phi_1, \dots, \phi_k, \boldsymbol{\gamma})',$$

where $\boldsymbol{\gamma} = (\gamma_{11}, \dots, \gamma_{r1}, \gamma_{22}, \dots, \gamma_{r2}, \dots, \gamma_{rr})$. Then the $\text{EA}_k[f]$ model is expressed as

$$\boldsymbol{F}(2\boldsymbol{\pi}^c) = \boldsymbol{X}\boldsymbol{\beta} = (\boldsymbol{x_0}, \boldsymbol{x_1}, \dots, \boldsymbol{x_k}, \boldsymbol{X_{k+1}})\boldsymbol{\beta},$$

where

$$\begin{aligned} \boldsymbol{x_0} &= (\boldsymbol{v_1}, \dots, \boldsymbol{v_r})', \\ \boldsymbol{x_l} &= \boldsymbol{J_r^l} \otimes \boldsymbol{1_r} - \boldsymbol{1_r} \otimes \boldsymbol{J_r^l} \quad (l = 1, \dots, k), \end{aligned}$$

and $\boldsymbol{X_{k+1}}$ is the $r^2 \times r(r+1)/2$ matrix of 1 or 0 elements determined from the structure of the $\text{EA}_k[f]$ model. Note that $\boldsymbol{1_s}$ is the $s \times 1$ vector of 1 element, $\boldsymbol{v_t}$ is the $1 \times r$ vector of 0 for the first t elements and the others are 1, $\boldsymbol{J_r^l} = (u_1^l, \dots, u_r^l)'$, and "$\otimes$" denotes the Kronecker product. $\boldsymbol{X_{k+1}}\boldsymbol{1_{r(r+1)/2}} = \boldsymbol{1_{r^2}}$ holds, and the $r^2 \times K$ matrix $\boldsymbol{X}$ is full column rank where $K = k+1+r(r+1)/2$.

We denote the linear space spanned by the columns of the matrix $\boldsymbol{X}$ by $S(\boldsymbol{X})$ with the dimension K. Let $\boldsymbol{U}$ be an $r^2 \times d_1$, where $d_1 = r(r-1)/2-k-1$, the full column rank matrix such that the linear space spanned by the column of $\boldsymbol{U}$, that is, $S(\boldsymbol{U})$ is the orthogonal complement of the space $S(\boldsymbol{X})$. Thus, $\boldsymbol{U}'\boldsymbol{X}$ is the $d_1 \times K$ zero matrix.

Let $\boldsymbol{h_1}(\boldsymbol{\pi})$ be a vector of functions defined by $\boldsymbol{h_1}(\boldsymbol{\pi}) = \boldsymbol{U}'\boldsymbol{F}(2\boldsymbol{\pi}^c)$. Additionally, let $\boldsymbol{h_2}(\boldsymbol{\pi})$ be a vector of functions defined by $\boldsymbol{h_2}(\boldsymbol{\pi}) = \boldsymbol{M}\boldsymbol{\pi}$ with

$$\boldsymbol{M} = \begin{pmatrix} (2\boldsymbol{x_0} - \boldsymbol{1}_{r^2} + \boldsymbol{w_1} + \boldsymbol{w_2} + \cdots + \boldsymbol{w_r})' \\ \boldsymbol{x_1'} \\ \vdots \\ \boldsymbol{x_k'} \end{pmatrix} : \text{the } d_2 \times r^2 \text{ matrix},$$

where $d_2 = k + 1$ and $\boldsymbol{w_i}$ $(i = 1, \dots, r)$ is the $r^2 \times 1$ vector, being one of column vectors in $\boldsymbol{X_{k+1}}$ corresponding γ_{ii}. We note that $\boldsymbol{M}'$ belongs to the space $S(\boldsymbol{X})$, namely, $S(\boldsymbol{M}') \subset S(\boldsymbol{X})$. The $\text{EA}_k[f]$ model is equivalent to the hypothesis $\boldsymbol{h_1}(\boldsymbol{\pi}) = \boldsymbol{0_{d_1}}$, where $\boldsymbol{0_{d_1}}$ is the $d_1 \times 1$ zero vector. Similarly, the

GM_k model is equivalent to the hypothesis $\boldsymbol{h_2}(\boldsymbol{\pi}) = \boldsymbol{0_{d_2}}$. From Theorem 3.1, the S model is equivalent to the hypothesis $\boldsymbol{h_3}(\boldsymbol{\pi}) = \boldsymbol{0_{d_3}}$, where $\boldsymbol{h_3} = (\boldsymbol{h_1'}, \boldsymbol{h_2'})'$ and $d_3 = d_1 + d_2 = r(r-1)/2$.

Let $\boldsymbol{\Sigma}(\boldsymbol{\pi}) = diag(\boldsymbol{\pi}) - \boldsymbol{\pi}\boldsymbol{\pi}'$, where $diag(\boldsymbol{\pi})$ denotes a diagonal matrix with the ith component of $\boldsymbol{\pi}$ as the ith diagonal component. Let $\boldsymbol{p}$ denote $\boldsymbol{\pi}$ with π_{ij} replaced by p_{ij}, where $p_{ij} = n_{ij}/n$. Then $\sqrt{n}(\boldsymbol{p} - \boldsymbol{\pi})$ has an asymptotic normal distribution with mean $\boldsymbol{0_{r^2}}$ and covariance matrix $\boldsymbol{\Sigma}(\boldsymbol{\pi})$. Using the delta method, $\sqrt{n}(\boldsymbol{h_3}(\boldsymbol{p}) - \boldsymbol{h_3}(\boldsymbol{\pi}))$ has an asymptotic normal distribution with mean $\boldsymbol{0_{d_3}}$ and covariance matrix

$$\boldsymbol{H_3}(\boldsymbol{\pi})\boldsymbol{\Sigma}(\boldsymbol{\pi})\boldsymbol{H_3'}(\boldsymbol{\pi}) = \begin{pmatrix} \boldsymbol{H_1}(\boldsymbol{\pi})\boldsymbol{\Sigma}(\boldsymbol{\pi})\boldsymbol{H_1'}(\boldsymbol{\pi}) & \boldsymbol{H_1}(\boldsymbol{\pi})\boldsymbol{\Sigma}(\boldsymbol{\pi})\boldsymbol{H_2'}(\boldsymbol{\pi}) \\ \boldsymbol{H_2}(\boldsymbol{\pi})\boldsymbol{\Sigma}(\boldsymbol{\pi})\boldsymbol{H_1'}(\boldsymbol{\pi}) & \boldsymbol{H_2}(\boldsymbol{\pi})\boldsymbol{\Sigma}(\boldsymbol{\pi})\boldsymbol{H_2'}(\boldsymbol{\pi}) \end{pmatrix},$$

where $\boldsymbol{H_s}(\boldsymbol{\pi}) = \partial \boldsymbol{h_s}(\boldsymbol{\pi})/\partial \boldsymbol{\pi}'$ $(s = 1, 2, 3)$. Then

$$\boldsymbol{H_1}(\boldsymbol{\pi}) = \boldsymbol{U}'(diag(\boldsymbol{a}) + \boldsymbol{J} diag(\boldsymbol{b})),$$

where

$$\boldsymbol{a} = (a_{11}, \dots, a_{1r}, a_{21}, \dots, a_{2r}, \dots, a_{r1}, \dots, a_{rr})',$$
$$\boldsymbol{b} = (b_{11}, \dots, b_{1r}, b_{21}, \dots, b_{2r}, \dots, b_{r1}, \dots, b_{rr})',$$

with

$$a_{ij} = \frac{\partial}{\partial \pi_{ij}} F\left(2\pi_{ij}^c\right), \quad b_{ij} = \frac{\partial}{\partial \pi_{ij}} F\left(2\pi_{ji}^c\right),$$

and $\boldsymbol{J}$ denotes the $r^2 \times r^2$ matrix such that

$$\boldsymbol{J}\boldsymbol{\pi} = (\pi_{11}, \dots, \pi_{r1}, \pi_{12}, \dots, \pi_{r2}, \dots, \pi_{1r}, \dots, \pi_{rr})'.$$

Additionally, we can obtain $\boldsymbol{H_2}(\boldsymbol{\pi}) = \boldsymbol{M}$. Under the hypothesis $\boldsymbol{h_3}(\boldsymbol{\pi}) = \boldsymbol{0_{d_3}}$,

$$\begin{aligned} \boldsymbol{H_1}(\boldsymbol{\pi})\boldsymbol{\Sigma}(\boldsymbol{\pi})\boldsymbol{H_2'}(\boldsymbol{\pi}) &= \boldsymbol{H_1}(\boldsymbol{\pi})(diag(\boldsymbol{\pi}) - \boldsymbol{\pi}\boldsymbol{\pi}')\boldsymbol{M}', \\ &= \boldsymbol{H_1}(\boldsymbol{\pi}) diag(\boldsymbol{\pi})\boldsymbol{M}', \end{aligned}$$

because $\boldsymbol{M}\boldsymbol{\pi}$ is the $d_2 \times 1$ zero vector. Then

$$\boldsymbol{H_1}(\boldsymbol{\pi}) diag(\boldsymbol{\pi}) = \frac{\bar{F}(1)}{2}\boldsymbol{U}'(\boldsymbol{I} - \boldsymbol{J}),$$

where $\bar{F}(u) = dF(u)/du$. We obtain $\boldsymbol{H_1}(\boldsymbol{\pi})\boldsymbol{\Sigma}(\boldsymbol{\pi})\boldsymbol{H_2'}(\boldsymbol{\pi})$ is the $d_1 \times d_2$ zero matrix since $\boldsymbol{J}\boldsymbol{M}' = -\boldsymbol{M}'$. Thus,

$$W_3 = W_1 + W_2,$$

where

$$W_s = n\boldsymbol{h_s}(\boldsymbol{p})'(\boldsymbol{H_s}(\boldsymbol{p})\boldsymbol{\Sigma}(\boldsymbol{p})\boldsymbol{H_s'}(\boldsymbol{p}))^{-1}\boldsymbol{h_s}(\boldsymbol{p}).$$

The Wald statistic W_s has an asymptotic chi-squared distribution with d_s degrees of freedom. That is, (i) W_1 is the Wald statistic for the $\mathrm{EA}_k[f]$ model. (ii) W_2 is that for the GM_k model. (iii) W_3 is that for the S model. The proof is completed from the asymptotic equivalence of the Wald statistic and the likelihood ratio statistic proved by Rao (1973). □

When $k = 0$, we obtain the following corollary.

Corollary 4.1. *Under the S model, the test statistic $G^2(S)$ is asymptotically equivalent to the sum of $G^2(EA_0[f])$ and $G^2(GS)$.*

§5. Example

The data in Table 1 from Bishop, Fienberg, and Holland (1975, p.210), are constructed from the occupational status of 2,391 father-son pairs in Denmark. The categories are ordinal from (1) to (5) (high to low). For a fixed k ($k = 0, 1, \ldots, r-1$), the PEA_k model is applied to the data in Table 1 with the integer score. Table 2 gives the value of the likelihood ratio statistic G^2 for each model.

The S model fits the data in Table 1 poorly, whereas the PEA_3, PEA_4, and GM_2 models fit well. We consider the hypothesis that the PEA_3 model holds assuming that the PEA_4 model holds, namely, $\alpha_4 = 0$. Using a test based on the difference between the likelihood ratio chi-square statistic, this hypothesis is accepted at the 0.05 significance level because the difference between two likelihood ratio chi-square values is 0.871 with 1 df. Therefore, the PEA_3 model is preferable to the PEA_4 model, that is, it is reasonable that the difference of the conditional probabilities is expressed as cubic than quartic with respect to $\{u_i\}$.

The values of MLEs of α_0, α_1, α_2, and α_3 for the PEA_3 model are 0.086, 4.131, -1.358, and 0.133, respectively. We estimate the difference between two conditional probabilities as $\hat{\pi}^c_{ij} - \hat{\pi}^c_{ji} = (0.086 + 4.131(i-j) - 1.358(i^2 - j^2) + 0.133(i^3 - j^3))/4$ for $i < j$. The estimated differences between two conditional probabilities $\hat{\pi}^c_{1j} - \hat{\pi}^c_{j1}$ ($j = 2, \ldots, 5$) are all negative. In this case, the conditional probability that the son's status belongs to (1) under condition that the son's status belongs to (1) and the father's status belongs to (j) or the son's status belongs to (j) and the father's status belongs to (1) is estimated to be higher than the conditional probability that the father's status belongs to (1) under same condition. By contrast, the estimated differences between two conditional probabilities $\hat{\pi}^c_{ij} - \hat{\pi}^c_{ji}$ ($i = 2, 3, 4; j > i$) are all positive. In this case, the conditional probability that the father's status belongs to (i) under condition that the father's status belongs to (i) and the son's status belongs to (j) or the father's status belongs to (j) and the son's status belongs to (i) is estimated to be higher than the conditional probability that the son's status belongs to (i) under same condition.

We are interested in inferring the reason for the poor fit of the S model. According to Theorem 3.1, the S model can be separated into the PEA_3 model and the GM_3 model. The PEA_3 model fits very well, but the GM_3 model fits very poorly. Hence, we deduce that the lack of structure of the GM_3 model is responsible for the poor fit of the S model.

§6. Discussion

We propose an asymmetry model based on the f-divergence, which includes various asymmetry models. We discuss the choice of f. The $\mathrm{EA}_k[f]$ model is concerned with conditional probabilities. Considering the interpretable model, we can easily interpret the PEA_k model because the PEA_k model can be expressed as the difference between π_{ij}^c and π_{ji}^c. However, it is difficult to interpret models based on power-divergence, except the ELS_k and PEA_k models. The PEA_k model is useful to analyze square contingency tables with ordinal categories.

Acknowledgments

We are greatly thankful to anonymous referee for helpful and valuable suggestions to further improve this paper.

References

[1] Agresti, A. (1983). A simple diagonals-parameter symmetry and quasi-symmetry model. *Statistics and Probability Letters*, **1**, 313–316.

[2] Aitchison, J. (1962). Large-sample restricted parametric tests. *Journal of the Royal Statistical Society*, Ser. B, **24**, 234–250.

[3] Bhapkar, V. P. (1966). A note on the equivalence of two test criteria for hypotheses in categorical data. *Journal of the American Statistical Association*, **61**, 228–235.

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

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

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

[7] Csiszár, I. and Shields, P. C. (2004). *Information Theory and Statistics: A tutorial.* Now Publishers, Hanover.

[8] Darroch, J. N. and Silvey, S. D. (1963). On testing more than one hypothesis. *The Annals of Mathematical Statistics*, **34**, 555–567.

[9] Kateri, M. and Agresti, A. (2007). A class of ordinal quasi-symmetry models for square contingency tables. *Statistics and Probability Letters*, **77**, 598–603.

[10] Kateri, M. and Papaioannou, T. (1997). Asymmetry models for contingency tables. *Journal of the American Statistical Association*, **92**, 1124–1131.

[11] Lang, J. B. and Agresti, A. (1994). Simultaneously modeling joint and marginal distributions of multivariate categorical responses. *Journal of the American Statistical Association*, **89**, 625–632.

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

[13] Rao, C. R. (1973). *Linear Statistical Inference and Its Applications*, 2nd edition. Wiley, New York.

[14] Read, C. B. (1977). Partitioning chi-square in contingency tables: A teaching approach. *Communications in Statistics-Theory and Methods*, **6**, 553–562.

[15] Read, T. R. C. and Cressie, N. A. C. (1988). *Goodness-of-Fit Statistics for Discrete Multivariate Data.* Springer-Verlag, New York.

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

[17] Tahata, K. (2019). Separation of symmetry for square tables with ordinal categorical data. *Japanese Journal of Statistics and Data Science*, (Available https://doi.org/10.1007/s42081-019-00066-8, [27 December, 2019]).

[18] Tahata, K., Naganawa, M., and Tomizawa, S. (2016). Extended linear asymmetry model and separation of symmetry for square contingency tables. *Journal of the Japan Statistical Society*, **46**, 189–202.

[19] Tomizawa, S. (1987). Decompositions for 2-ratios-parameter symmetry model in square contingency tables with ordered categories. *Biometrical Journal*, **29**, 45–55.

[20] Tomizawa, S. and Tahata, K. (2007). The analysis of symmetry and asymmetry: Orthogonality of decomposition of symmetry into quasi-symmetry and marginal symmetry for multi-way tables. *Journal de la Société Française de Statistique*, **148**, 3–36.

Table 1: Occupational status for father-son pairs; taken directly from Bishop *et al.* (1975, p.210)

Father's status	Son's status (1)	(2)	(3)	(4)	(5)	Total
(1)	18	17	16	4	2	57
	[a]18.000	15.861	15.704	5.499	3.630	
	[b]18.000	14.795	15.524	5.300	3.549	
(2)	24	105	109	59	21	318
	25.139	105.000	101.616	62.867	16.751	
	26.205	105.000	105.614	63.807	17.192	
(3)	23	84	289	217	95	708
	23.296	91.384	289.000	221.985	92.134	
	23.476	87.386	289.000	217.015	91.121	
(4)	8	49	175	348	198	778
	6.501	45.133	170.015	348.000	201.974	
	6.700	44.193	174.985	348.000	204.300	
(5)	6	8	69	201	246	530
	4.370	12.249	71.866	197.026	246.000	
	4.451	11.808	72.879	194.700	246.000	
Total	79	263	658	829	562	2391

Notes: [a]Estimated expected frequencies from the PEA_3 model.
[b]Estimated expected frequencies from the PEA_4 model.

Table 2: Likelihood ratio chi-square values G^2 for each model applied to the data in Table 1

Model	Degree of freedom	G^2
S	10	24.802*
PEA_0	9	18.819*
PEA_1	8	18.682*
PEA_2	7	17.227*
PEA_3	6	7.337
PEA_4	5	6.466
GS	1	5.983*
GM_1	2	6.120*
GM_2	3	7.576
GM_3	4	17.465*
GM_4	5	18.336*

Notes: *Significant at the 0.05 level

Kengo Fujisawa
Department of Information Sciences, Faculty of Science and Technology, Tokyo University of Science
Noda City, Chiba, 278-8510, Japan

Kouji Tahata
Department of Information Sciences, Faculty of Science and Technology, Tokyo University of Science
Noda City, Chiba, 278-8510, Japan
E-mail: kouji_tahata@rs.tus.ac.jp