A Bivariate Index Vector to Measure Departure from Quasi-symmetry for Ordinal Square Contingency Tables

Shuji Ando
Tokyo University of Science

Abstract

This study proposes a bivariate index vector to concurrently analyze both the degree and direction of departure from the quasi-symmetry (QS) model for ordinal square contingency tables. The QS model and extended QS (EQS) models identify the symmetry and asymmetry between the probabilities of normal circulation and reverse circulation when the order exists for arbitrary three categories. The asymmetry parameter of the EQS model implies the degree of departure from the QS model; the EQS model is equivalent to the QS model when the asymmetry parameter equals to one. The structure of the EQS model differs depending on whether the asymmetry parameter approaches zero or infinity. Thus, the asymmetry parameter of the EQS model also implies the direction of departure from the QS model. The proposed bivariate index vector is constructed by combining existing and original sub-indexes that represent the degree of departure from the QS model and its direction. These sub-indexes are expressed as functions of the asymmetry parameter under the EQS model. We construct an estimator of the proposed bivariate index vector and an approximate confidence region for the proposed bivariate index vector. Using real data, we show that the proposed bivariate index vector is important to compare degrees of departure from the QS model for plural data sets.

Keywords: asymmetry, circulation, comparison, direction, ordered categorical data, symmetry.

1. Introduction

Consider an \( R \times R \) square contingency table with the same row and column ordinal classifications. We denote \( \pi_{ij} \) as the probability that an observation will fall in \( (i, j) \)th cell of the table \( (i, j = 1, \ldots, R) \). The quasi-symmetry (QS) model proposed by Caussinus (1965) indicates the symmetry between the probabilities of the normal circulation \( \pi_{ij} \pi_{jk} \pi_{ki} \) and the probabilities of the reverse circulation \( \pi_{kj} \pi_{ji} \pi_{ik} \) for arbitrary \( i < j < k \). The extended QS (EQS) model proposed by Tomizawa (1984) indicates the asymmetry between them. The asymmetry parameter of the EQS model implies the degree of departure from the QS model because the EQS model is equivalent to the QS model when the asymmetry parameter equals one. The structure of the EQS model differs depending on whether the parameter approaches zero or infinity. Thus, the asymmetry parameter of the EQS model also implies the direction of departure from the QS model.

When the QS model does not hold for a single data set, we may apply the EQS model or
measure the degree of departure from the QS model. Test statistics (e.g., Pearson’s chi-squared statistic or likelihood ratio statistic) cannot judge how much the degree of departure from the QS model is toward the maximum departure. On the contrary, when the QS model does not hold for plural data sets, we could compare the degrees of departure from the QS model. Test statistics, however, may not be valuable for comparing the degrees of departure from the model in plural data sets because the value of test statistics depends on the sample size.

Tahata, Kozai, and Tomizawa (2014) proposed an index to measure the degree of departure from the QS model (see Section 3 for the details of this index). This index is important to measure and compare the degree of departure from the QS model as the range of this index is zero to one, and the value of this index does not depend on the sample size. This index, however, cannot distinguish directions of departure from the QS model because the value of this index approaches one as the asymmetry parameter approaches zero or infinity under the EQS model. To tackle this issue, we may consider a sequential estimation procedure, for example, estimate the index, and then decide the direction though estimating the asymmetry parameter of EQS model. This procedure, however, can only be used when the EQS model holds. We are interested in constructing an index that can distinguish directions of departure from the QS model even when the EQS model does not hold.

For measuring the degree of departure from the symmetry (S) model, Tomizawa, Miyamoto, and Hatanaka (2001) proposed an index that can judge whether or not the S model holds, and Tahata, Miyazawa, and Tomizawa (2010) proposed an index that can distinguish directions of departure from the S model. The index of Tahata et al. (2010), however, cannot judge whether or not the S model holds. To concurrently analyze both the degree and direction of departure from the S model, Ando, Tahata, and Tomizawa (2017) proposed a bivariate index vector with the above indexes as elements.

This study proposes (1) an index that can distinguish directions of departure from the QS model, and (2) a bivariate index vector that can concurrently analyze both the degree and direction of departure from the QS model in a manner similar to Ando et al. (2017). The proposed bivariate index vector can be visually comparing the degrees of departure from the QS model in plural data sets using confidence region. This is useful to produce results easier to interpret than the existing index.

This paper is organized as follows. Section 2 introduces the QS and EQS models. Section 3 defines the proposed bivariate index vector. Section 4 derives an estimator and approximate confidence region for the proposed bivariate index vector. Section 5 demonstrates the utility of the proposed index vector using application to real data sets. Section 6 closes with concluding remarks.

2. Quasi-symmetry model

Caussinus (1965) considered the QS model defined by

$$\pi_{ij} = \alpha_i \beta_j \psi_{ij} \quad \text{for } i, j = 1, \ldots, R,$$

where $\psi_{ij} = \psi_{ji}$. The QS model with $\{\alpha_i = \beta_i\}$ is equivalent to the S model proposed by Bowker (1948). The QS model is also expressed as

$$\pi_{ij} \pi_{jk} \pi_{ki} = \pi_{kj} \pi_{ji} \pi_{ik} \quad \text{for } i < j < k,$$

see, for example, Tomizawa (1984), Tahata et al. (2014), and Altun (2019). The QS model indicates that the probabilities of the normal circulation $\pi_{ij} \pi_{jk} \pi_{ki}$ are equal to the probabilities of the reverse circulation $\pi_{kj} \pi_{ji} \pi_{ik}$ for arbitrary $i < j < k$.

Tomizawa (1984) considered the EQS model defined by

$$\pi_{ij} \pi_{jk} \pi_{ki} = \gamma \pi_{kj} \pi_{ji} \pi_{ik} \quad \text{for } i < j < k.$$
The EQS model indicates that the probabilities of the normal circulation $\pi_{ij}$ $\pi_{jk}$ $\pi_{ki}$ are $\gamma$ times higher than the probabilities of the reverse circulation $\pi_{kj}$ $\pi_{ji}$ $\pi_{ik}$ for arbitrary $i<j<k$.

The number of degrees of freedom for the QS and EQS models are $(R-1)(R-2)/2$ and $R(R-3)/2$.

The parameter $\gamma$ of the EQS model implies the degree of departure from the QS model because the EQS model with $\gamma = 1$ is equivalent to the QS model. The $\gamma$ also implies the direction of departure from the QS model because the structure of the EQS model differs depending on whether the $\gamma$ approaches zero or infinity.

The QS model is a typical model among many models for the square contingency table. The issues related to the QS model have been analyzed in many studies, for example, Kateri and Papaioannou (1997), Kateri and Agresti (2007), Saigusa, Tahata, and Tomizawa (2015), Kateri, Gottard, and Tarantola (2017), Altun (2019), Tahata (2019), and the references therein.

3. Proposed methods

In this section, we propose a bivariate index vector that can concurrently analyze both the degree and direction of departure from the QS model.

3.1. Existing index of quasi-symmetry model

Let $A_{ijk} = \pi_{ij}\pi_{jk}\pi_{ki}$ and $A_{kji} = \pi_{kj}\pi_{ji}\pi_{ik}$ for $i<j<k$.

Assume that $A_{ijk} + A_{kji} > 0$ for $i<j<k$, Tahata et al. (2014) proposed the index $\phi$ defined by

$$\phi = \frac{1}{\log 2} \sum_{i<j<k} \left[ B_{ijk} \log \left( \frac{B_{ijk}}{C_{ijk}} \right) + B_{kji} \log \left( \frac{B_{kji}}{C_{ijk}} \right) \right],$$

where

$$B_{ijk} = \frac{A_{ijk}}{\Delta}, \quad B_{kji} = \frac{A_{kji}}{\Delta}, \quad C_{ijk} = \frac{B_{ijk} + B_{kji}}{2} \quad \text{and} \quad \Delta = \sum_{i<j<k} (A_{ijk} + A_{kji}).$$

The $\phi$ has the following characteristics: (1) the range of $\phi$ is zero to one, (2) $\phi = 0$ if and only if the QS model holds, and (3) $\phi = 1$ if and only if there is maximum departure from the QS model. The maximum departure from the QS model implies $A_{ijk} = 0$ or $A_{kji} = 0$ for $i<j<k$.

Under the EQS model, the $\phi$ is also expressed as

$$\phi = \frac{1}{\log 2} \left[ \frac{\gamma}{\gamma + 1} \log \left( \frac{2\gamma}{\gamma + 1} \right) + \frac{1}{\gamma + 1} \log \left( \frac{2}{\gamma + 1} \right) \right].$$

We see that if the $\gamma$ approaches zero or infinity then the $\phi$ approaches both one. We are interested in considering the index that can take a difference value when the $\gamma$ approaches zero and infinity because the structure differs depending on whether the $\gamma$ approaches zero and infinity.

3.2. Directional index of quasi-symmetry model

We propose a new directional index that can distinguish directions of departure from the QS model. This index can take a difference value when the $\gamma$ approaches zero or infinity under the EQS model.

Assume that $A_{ijk} + A_{kji} > 0$ for $i<j<k$, we define the directional index $\varphi$ as

$$\varphi = \frac{4}{\pi} \sum_{i<j<k} \left( B_{ijk} + B_{kji} \right) \left( \theta_{ijk} - \frac{\pi}{4} \right),$$
where
\[ \theta_{ijk} = \arccos \left( \frac{A_{ijk}}{\sqrt{(A_{ijk})^2 + (A_{kji})^2}} \right). \]

The \( \varphi \) has the following characteristics: (1) the range of \( \varphi \) is minus one to one, (2) \( \varphi = -1 \) if and only if \( A_{kji} \) equals to zero for \( i < j < k \) (say, maximum reverse circulation), (3) \( \varphi = 1 \) if and only if \( A_{ijk} \) equals to zero for \( i < j < k \) (say, maximum normal circulation), and (4) if the QS model holds then \( \varphi = 0 \), but the converse does not hold. We refer to the structure of \( \varphi = 0 \) as average circulation. From the above property (4), we note that the \( \varphi \) concurrently cannot analyze both the degree and direction of departure from the QS model.

Under the EQS model, the \( \varphi \) is also expressed as
\[ \varphi = \frac{4}{\pi} \arccos \left( \frac{\gamma}{\sqrt{1 + \gamma^2}} \right) - 1. \]

We see that if the \( \gamma \) approaches infinity then the \( \varphi \) approaches minus one, and if the \( \gamma \) approaches zero then the \( \varphi \) approaches one. The \( \varphi \) can take a difference value when the \( \gamma \) approaches zero or infinity.

### 3.3. Bivariate index vector of the quasi-symmetry model

To concurrently analyze both the degree and direction of departure from the QS model, we propose the bivariate index vector as \( \Phi = (\phi, \varphi)' \). Note that the symbol “ ‘ ” denotes the transpose.

The proposed bivariate index vector \( \Phi \) satisfies the following properties.

1. \( \Phi = (0, 0)' \) if and only if the QS model holds;
2. \( \Phi = (1, -1)' \) if and only if the structure of maximum reverse circulation exists;
3. \( \Phi = (1, 1)' \) if and only if the structure of maximum normal circulation exists.

From the above properties, the bivariate index vector \( \Phi \) concurrently can analyze both the degree and direction of departure from the QS model.

### 4. Approximate confidence region for the bivariate index vector

Let \( n_{ij} \) denote the observed frequency in the \((i, j)\)th cell of the table \((i, j = 1, \ldots, R)\). Assume that a multinomial distribution applies to the \( R \times R \) table. Let \( p_{ij} = n_{ij}/N \) \((i, j = 1, \ldots, R)\) and \( N = \sum \sum n_{ij} \). We estimate \( \Phi \) by \( \hat{\Phi} = (\hat{\phi}, \hat{\varphi})' \), where \( \hat{\phi} \) and \( \hat{\varphi} \) are given by \( \phi \) and \( \varphi \) with \( \{\pi_{ij}\} \) being replaced by \( \{p_{ij}\} \), respectively. Using the delta method (see, e.g., Agresti 2013, p. 591), \( \sqrt{N}(\Phi - \Phi) \) asymptotically (as \( N \to \infty \)) has a bivariate normal distribution with zero mean and covariance matrix \( \Sigma[\hat{\Phi}] \),

\[
\Sigma[\hat{\Phi}] = \frac{1}{N^2} \sum_{s=1}^{R-1} \sum_{t=s+1}^{R} \begin{pmatrix}
(p_{st} (\alpha_{st})^2 + p_{ts} (\alpha_{ts})^2) & p_{st} \alpha_{st} \beta_{st} + p_{ts} \alpha_{ts} \beta_{ts} \\
p_{st} \alpha_{st} \beta_{st} + p_{ts} \alpha_{ts} \beta_{ts} & (\beta_{st})^2 + p_{ts} (\beta_{ts})^2
\end{pmatrix},
\]
where

\[
\alpha_{st} = \frac{1}{\log 2} \sum_{i<j<k} \left[ \frac{\partial A_{ijk}}{\partial \pi_{st}} \log \left( \frac{2A_{ijk}}{A_{ijk} + A_{kji}} \right) + \frac{\partial A_{kji}}{\partial \pi_{st}} \log \left( \frac{2A_{kji}}{A_{ijk} + A_{kji}} \right) \right]
\]

\[
\alpha_{ts} = \frac{1}{\log 2} \sum_{i<j<k} \left[ \frac{\partial A_{ijk}}{\partial \pi_{ts}} \log \left( \frac{2A_{ijk}}{A_{ijk} + A_{kji}} \right) + \frac{\partial A_{kji}}{\partial \pi_{ts}} \log \left( \frac{2A_{kji}}{A_{ijk} + A_{kji}} \right) \right]
\]

\[
\beta_{st} = \frac{4}{\pi} \sum_{i<j<k} \left[ \left( \frac{\partial A_{ijk}}{\partial \pi_{st}} + \frac{\partial A_{kji}}{\partial \pi_{st}} \right) \theta_{ijk} + (A_{ijk} + A_{kji}) \frac{\partial \theta_{ijk}}{\partial \pi_{st}} \right]
\]

\[
\beta_{ts} = \frac{4}{\pi} \sum_{i<j<k} \left[ \left( \frac{\partial A_{ijk}}{\partial \pi_{ts}} + \frac{\partial A_{kji}}{\partial \pi_{ts}} \right) \theta_{ijk} + (A_{ijk} + A_{kji}) \frac{\partial \theta_{ijk}}{\partial \pi_{ts}} \right]
\]

with

\[
\frac{\partial \theta_{ijk}}{\partial \pi_{st}} = \frac{\partial A_{kji}}{\partial \pi_{st}} A_{ijk} - \frac{\partial A_{ijk}}{\partial \pi_{st}} A_{kji}
\]

\[
\frac{\partial \theta_{ijk}}{\partial \pi_{ts}} = \frac{\partial A_{kji}}{\partial \pi_{ts}} A_{ijk} - \frac{\partial A_{ijk}}{\partial \pi_{ts}} A_{kji}
\]

\[
\frac{\partial A_{ijk}}{\partial \pi_{st}} = I(i = s, j = t)\pi_{jk}\pi_{ki} + I(j = s, k = t)\pi_{ij}\pi_{ki},
\]

\[
\frac{\partial A_{ijk}}{\partial \pi_{ts}} = I(k = t, i = s)\pi_{ij}\pi_{jk},
\]

\[
\frac{\partial A_{kji}}{\partial \pi_{st}} = I(i = s, k = t)\pi_{kj}\pi_{ji},
\]

\[
\frac{\partial A_{kji}}{\partial \pi_{ts}} = I(k = t, j = s)\pi_{ji}\pi_{ik} + I(j = t, i = s)\pi_{kj}\pi_{ik},
\]

and \(I(\cdot)\) is the indicator function. We note that the asymptotic normal distribution of \(\sqrt{N}(\hat{\Phi} - \Phi)\) is not applicable when \(\phi = 0\) and \(\phi = 1\) because the variance of \(\hat{\phi}\) equals zero, and also not applicable when \(\phi = -1\) and \(\phi = 1\) because the variance of \(\hat{\phi}\) equals zero.

Let \(\Sigma[\hat{\Phi}]\) denote \(\Sigma[\hat{\Phi}]\) with \(\{\pi_{ij}\}\) being replaced by \(\{\hat{p}_{ij}\}\). An approximate 100(1 – \(\alpha\)) percent confidence region for \(\Phi\) is obtained by

\[
N(\hat{\Phi} - \Phi)'\left(\Sigma[\hat{\Phi}]\right)^{-1}(\hat{\Phi} - \Phi) \leq \chi^2_{(1-\alpha;2)},
\]

where \(\chi^2_{(1-\alpha;2)}\) is the 1 – \(\alpha\) quantile of the chi-square distribution with two degrees of freedom.

Consider comparing the degrees of departure from the QS model between independent Tables A and B (with sample sizes \(N_A\) and \(N_B\), respectively). For Tables A and B, the \(\Phi\) are denoted by \(\Phi_A\) and \(\Phi_B\), respectively. We also denote estimators of \(\Phi_A\) and \(\Phi_B\) by \(\hat{\Phi}_A\) and
When $N_A$ and $N_B$ are large, $\hat{\Phi}_A - \hat{\Phi}_B - (\Phi_A - \Phi_B)$ approximately has a bivariate normal distribution with zero mean and covariance matrix $\mathbf{V}$, where

$$\mathbf{V} = \frac{1}{N_A} \Sigma[\hat{\Phi}_A] + \frac{1}{N_B} \Sigma[\hat{\Phi}_B].$$

Therefore, an approximate 100$(1 - \alpha)$ percent confidence region for $\Phi_A - \Phi_B$ is obtained by

$$(\hat{\Phi}_A - \hat{\Phi}_B - (\Phi_A - \Phi_B))^T \hat{\mathbf{V}}^{-1} (\hat{\Phi}_A - \hat{\Phi}_B - (\Phi_A - \Phi_B)) \leq \chi^2_{1 - \alpha; 2},$$

where $\hat{\mathbf{V}}$ is given by $\mathbf{V}$ with $\Sigma[\hat{\Phi}_A]$ and $\Sigma[\hat{\Phi}_B]$ replaced by $\hat{\Sigma}[\phi_A]$ and $\hat{\Sigma}[\phi_B]$, respectively.

We note that (i) if the confidence region for $\Phi_A - \Phi_B$ is included within the first quadrant, then the degree of departure from the QS model in Table A is greater than that in Table B, and the degree of normal circulation in Table A is greater than that in Table B; (ii) if the confidence region is included within the second quadrant, then the degree of departure from the QS model in Table A is greater than that in Table B, and the degree of reverse circulation in Table A is smaller than that in Table B; (iii) if the confidence region is included within the third quadrant, then the degree of departure from the QS model in Table A is smaller than that in Table B, and the degree of reverse circulation in Table A is smaller than that in Table B; and (iv) if the confidence region is included within the fourth quadrant, then the degree of departure from the QS model in Table A is smaller than that in Table B, and the degree of normal circulation in Table A is greater than that in Table B.

### 5. Application to data

#### 5.1. Application to artificial data

We point out that the bivariate index vector $\Phi$ is useful for visually comparing the degrees of departure from the QS model. For instance, we consider the three artificial data sets in Tables 1a, 1b, and 1c.

<table>
<thead>
<tr>
<th></th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>(a) Normal circulation</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(1) 1000</td>
<td>50</td>
<td>200</td>
<td>200</td>
<td>1450</td>
<td></td>
</tr>
<tr>
<td>(2) 200</td>
<td>1000</td>
<td>200</td>
<td>200</td>
<td>1600</td>
<td></td>
</tr>
<tr>
<td>(3) 200</td>
<td>50</td>
<td>1000</td>
<td>50</td>
<td>1300</td>
<td></td>
</tr>
<tr>
<td>(4) 50</td>
<td>200</td>
<td>200</td>
<td>1000</td>
<td>1450</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>1450</td>
<td>1300</td>
<td>1600</td>
<td>1450</td>
<td>5800</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>(b) Reverse circulation</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(1) 1000</td>
<td>200</td>
<td>200</td>
<td>50</td>
<td>1450</td>
<td></td>
</tr>
<tr>
<td>(2) 50</td>
<td>1000</td>
<td>50</td>
<td>200</td>
<td>1300</td>
<td></td>
</tr>
<tr>
<td>(3) 200</td>
<td>200</td>
<td>1000</td>
<td>200</td>
<td>1600</td>
<td></td>
</tr>
<tr>
<td>(4) 200</td>
<td>200</td>
<td>50</td>
<td>1000</td>
<td>1450</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>1450</td>
<td>1600</td>
<td>1300</td>
<td>1450</td>
<td>5800</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>(c) Average circulation</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(1) 1000</td>
<td>50</td>
<td>200</td>
<td>50</td>
<td>1300</td>
<td></td>
</tr>
<tr>
<td>(2) 200</td>
<td>1000</td>
<td>200</td>
<td>50</td>
<td>1450</td>
<td></td>
</tr>
<tr>
<td>(3) 200</td>
<td>200</td>
<td>1000</td>
<td>200</td>
<td>1600</td>
<td></td>
</tr>
<tr>
<td>(4) 200</td>
<td>200</td>
<td>50</td>
<td>1000</td>
<td>1450</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>1600</td>
<td>1450</td>
<td>1450</td>
<td>1300</td>
<td>5800</td>
</tr>
</tbody>
</table>
From Tables 2a, 2b, and 2c, we see that all of the values \( \hat{\phi} \) for Tables 1a, 1b, and 1c are equal, but the data sets in Tables 1a, 1b, and 1c have different structures for the direction of the departure from the QS model. The value of \( \hat{\varphi} \) for Table 1c equals zero, but the data set in Table 1c does not have the structure of the QS model. From Figure 1, we see that the proposed bivariate index vector \( \Phi \) can concurrently analyze the degree and direction of departure from the QS model. Thus, we see that (i) the data set in Table 1a tends to have the structure of normal circulation; (ii) the data set in Table 1b tends to have the structure of reverse circulation; and (iii) the data set in Table 1c tends to have the structure of average circulation.

Table 2: Estimates of \( \phi \) and \( \varphi \), approximate standard errors for \( \hat{\phi} \) and \( \hat{\varphi} \), and approximate 95% confidence intervals for \( \phi \) and \( \varphi \), applied to Tables 1a, 1b and 1c.

<table>
<thead>
<tr>
<th></th>
<th>Estimated index</th>
<th>Standard error</th>
<th>Confidence interval</th>
</tr>
</thead>
<tbody>
<tr>
<td>(a) For Table 1a</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( \phi )</td>
<td>0.461</td>
<td>0.042</td>
<td>(0.378, 0.543)</td>
</tr>
<tr>
<td>( \varphi )</td>
<td>0.626</td>
<td>0.050</td>
<td>(0.528, 0.724)</td>
</tr>
<tr>
<td>(b) For Table 1b</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( \phi )</td>
<td>0.461</td>
<td>0.042</td>
<td>(0.378, 0.543)</td>
</tr>
<tr>
<td>( \varphi )</td>
<td>-0.626</td>
<td>0.050</td>
<td>(-0.724, -0.528)</td>
</tr>
<tr>
<td>(c) For Table 1c</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( \phi )</td>
<td>0.461</td>
<td>0.042</td>
<td>(0.379, 0.542)</td>
</tr>
<tr>
<td>( \varphi )</td>
<td>0.000</td>
<td>0.068</td>
<td>(-0.132, 0.132)</td>
</tr>
</tbody>
</table>

5.2. Application to real data

We used the cross-classification of father’s and son’s occupational social class data in Japan. We took the two data sets from Hashimoto (2003, p. 142) and Tominaga (1979, p. 132) summarized in Tables 3 and 4. These data sets were examined in 1955 and 1975.

Table 3: Cross-classification of the Japanese father’s and son’s occupational status categories in 1955; taken from Hashimoto (2003, p. 142). Status is (1) capitalist, (2) new middle, (3) working, (4) self-employed, and (5) farming.

<table>
<thead>
<tr>
<th>Father’s status</th>
<th>Son’s status</th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
<th>(5)</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>(1)</td>
<td>39</td>
<td>39</td>
<td>39</td>
<td>57</td>
<td>23</td>
<td>197</td>
<td></td>
</tr>
<tr>
<td>(2)</td>
<td>12</td>
<td>78</td>
<td>23</td>
<td>23</td>
<td>37</td>
<td>173</td>
<td></td>
</tr>
<tr>
<td>(3)</td>
<td>6</td>
<td>16</td>
<td>78</td>
<td>23</td>
<td>20</td>
<td>143</td>
<td></td>
</tr>
<tr>
<td>(4)</td>
<td>18</td>
<td>80</td>
<td>79</td>
<td>126</td>
<td>31</td>
<td>334</td>
<td></td>
</tr>
<tr>
<td>(5)</td>
<td>28</td>
<td>106</td>
<td>136</td>
<td>122</td>
<td>628</td>
<td>1020</td>
<td></td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>103</strong></td>
<td><strong>319</strong></td>
<td><strong>355</strong></td>
<td><strong>351</strong></td>
<td><strong>739</strong></td>
<td><strong>1867</strong></td>
<td></td>
</tr>
</tbody>
</table>

First, we tested the goodness-of-fit of the QS model for Tables 3 and 4. The QS model is a poor fit for these data sets because the values of likelihood ratio statistic for Tables 3 and 4 are 22.13 and 8.30, respectively. The EQS model is also a poor fit for these data sets because its values are 13.59 and 8.28. From the above results, we cannot judge how much the degree of departure from the QS model is toward the maximum reverse or normal circulation. Furthermore, since these data sets differ in sample size and number of categories, it is not appropriate to compare the degrees of departure from the QS model using the values of likelihood ratio statistic.

Second, we measured the degree of departure from the QS model using the confidence region for \( \Phi \). For Tables 3 and 4, estimates of \( \Phi \) are
A bivariate index for measuring quasi-symmetry

Figure 1: Approximate 95% confidence regions for the bivariate index vector $\Phi = (\phi, \varphi)'$, applied to Tables 1a, 1b, and 1c.

Table 4: Cross-classification of the Japanese father’s and son’s occupational status categories in 1975; taken from Tominaga (1979, p. 132). Status is (1) professional and managers, (2) clerical and sales, (3) skilled manual and semiskilled manual, and (4) unskilled manual and farmers.

<table>
<thead>
<tr>
<th>Father’s status</th>
<th>Son’s status</th>
<th></th>
<th></th>
<th></th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(1)</td>
<td>(2)</td>
<td>(3)</td>
<td>(4)</td>
<td></td>
</tr>
<tr>
<td>(1)</td>
<td>127</td>
<td>101</td>
<td>54</td>
<td>12</td>
<td>294</td>
</tr>
<tr>
<td>(2)</td>
<td>86</td>
<td>207</td>
<td>125</td>
<td>13</td>
<td>431</td>
</tr>
<tr>
<td>(3)</td>
<td>78</td>
<td>124</td>
<td>310</td>
<td>24</td>
<td>536</td>
</tr>
<tr>
<td>(4)</td>
<td>109</td>
<td>206</td>
<td>437</td>
<td>325</td>
<td>1077</td>
</tr>
<tr>
<td>Total</td>
<td>400</td>
<td>638</td>
<td>926</td>
<td>374</td>
<td>2338</td>
</tr>
</tbody>
</table>
\( \hat{\Phi}_3 = \begin{pmatrix} 0.156 \\ 0.401 \end{pmatrix} \) and \( \hat{\Phi}_4 = \begin{pmatrix} 0.050 \\ 0.001 \end{pmatrix} \), respectively, and estimates of \( \Sigma[\hat{\Phi}] \) are

\[
\Sigma[\hat{\Phi}_3] = \begin{pmatrix} 6.777 & 8.176 \\ 8.176 & 16.715 \end{pmatrix} \quad \text{and} \quad \Sigma[\hat{\Phi}_4] = \begin{pmatrix} 2.735 & 0.588 \\ 0.588 & 35.107 \end{pmatrix},
\]

respectively. From Figure 2, we can see that the data sets in Tables 3 and 4 tend to have the structure of normal circulation and average circulation, respectively.

Finally, we compared the degrees of departure from the QS model in Tables 3 and 4 using the confidence region for \( \Phi_3 - \Phi_4 \). From Figure 3, the confidence region for \( \Phi_3 - \Phi_4 \) is included within both the first and fourth quadrants. We can infer that the degree of normal circulation in Table 3 is greater than that in Table 4, although we cannot conclude that the degree of departure from the QS model in Table 3 is greater than that in Table 4.

Figure 2: Approximate 95\% confidence regions for the bivariate index vector \( \Phi = (\phi, \varphi)' \), applied to Tables 3 and 4.
Figure 3: An approximate 95% confidence region for the bivariate index vector $\Phi_3 - \Phi_4$. 
6. Concluding remarks

For the analysis of square contingency tables with ordered categories, the index $\phi$ was proposed to measure the degree of departure from the QS model based on the probabilities for the normal circulation $\pi_{ij}\pi_{jk}\pi_{ki}$ and the probabilities for the reverse circulation $\pi_{kj}\pi_{ji}\pi_{ik}$ for arbitrary $i < j < k$. The index $\phi$ can judge whether or not the QS model holds, but cannot distinguish the direction of departure from the QS model (i.e., maximum reverse circulation or maximum normal circulation). This study proposed the new index, $\varphi$, to distinguish the directions of departure from the QS model. The index $\varphi$ can distinguish the directions of departure from the QS model, but cannot judge whether or not the QS model holds. To tackle these issues, this study also proposed the bivariate index vector $\Phi = (\phi, \varphi)'$ in order to concurrently analyze both the degree and direction of departure from the QS model. We derived the approximate confidence region for the bivariate index vector $\Phi$. The utility of the bivariate index vector $\Phi$ was verified by being applied to real data. The proposed bivariate index vector $\Phi$ can be visually comparing the degree of departure from the QS model in plural tables. This is useful to produce results easier to interpret than the existing index.

The indexes $\phi$ and $\varphi$ can measure the degree of departure from the QS model corresponding to the EQS model since these indexes are expressed as the function of $\gamma$ under the EQS model. On the contrary, Kateri et al. (2017) proposed the conditional QS (CQS) model as another extension of the QS model. The reader may be interested in measuring the degree of departure from the QS model corresponding to the CQS model, which is a matter to be considered in future studies.

Acknowledgements

The author would like to thank the anonymous reviewers and the editors for careful reading and comments to improve this paper.

References


**Affiliation:**

Shuji Ando  
Department of Information and Computer Technology  
Faculty of Engineering  
Tokyo University of Science  
Katsushika-ku, Tokyo, Japan  
E-mail: shuji.ando@rs.tus.ac.jp