Measure of Departure from Extended Bradley-Terry Model for Paired Comparisons

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@is.noda.tus.ac.jp

Keigo Kozai

Department of Information Sciences, Faculty of Science and Technology, Tokyo University of Science

Noda City, Chiba 278-8510, Japan

E-mail: keigo14@hotmail.co.jp

Sadao Tomizawa

Department of Information Sciences, Faculty of Science and Technology, Tokyo University of Science

Noda City, Chiba 278-8510, Japan

E-mail: tomizawa@is.noda.tus.ac.jp

Received: February 20, 2012 Accepted: March 2, 2012 Published: May 1, 2012

doi:10.5539/ijsp.v1n1p72 URL: http://dx.doi.org/10.5539/ijsp.v1n1p72

Abstract

For paired comparisons, we propose a measure to represent the degree of departure from the extended Bradley-Terry model. The measure is expressed by using the Kullback-Leibler information and it ranges between 0 and 1. The measure is applied to the win-loss standings of professional baseball league in Japan.

Keywords: Extended Bradley-Terry model, Kullback-Leibler information, Power-divergence

1. Introduction

Consider the athletic competitions with the outcome for the play of any two teams of R teams, namely a set of data from $R(R-1)/2$ paired comparison. Let π_{ij} for $i \neq j$ denote the probability that team i defeats team j when team i plays team j. Note that $\pi_{ji} = 1 - \pi_{ij}$ for $i < j$; that is, a tie cannot occur.

The Bradley-Terry (BT) model is defined by

$$\pi_{ij} = \frac{\delta_i}{\delta_i + \delta_j} \quad \text{for} \quad i \neq j.$$

This may be expressed by

$$G_{ijk} = G_{kji} \quad \text{for} \quad i < j < k,$$

where

$$G_{ijk} = \pi_{ij}\pi_{jk}\pi_{ki}, \quad G_{kji} = \pi_{kj}\pi_{ji}\pi_{ik};$$

see Bradley and Terry (1952), and Tahata, Miyamoto and Tomizawa (2004).

Assume that R teams are arranged in an order, for example, in order of ranking. The extended Bradley-Terry (EBT) model is defined by

$$\pi_{ij} = \frac{\gamma\delta_i}{\gamma\delta_i + \delta_j} \quad \text{for} \quad i < j.$$

This may be expressed as

$$G_{ijk} = \gamma G_{kji} \quad \text{for} \quad i < j < k;$$

see Davidson and Beaver (1977), and Agresti (1990, p. 373). A special case of this model obtained by putting $\gamma = 1$ is the BT model. This model indicates that for the plays of any two teams of teams i, j and k, the probability that team i defeats team j, team j defeats team k, and team k defeats team i, is γ times higher than the probability that k defeats j, j defeats i, and i defeats k.

For square tables with *nominal* categories, in which cells on the main diagonal are empty, Tahata et al. (2004) proposed the measures to represent the degree of departure from the BT model. We are interested in considering a measure which represents the degree of departure from the EBT model for square tables with *ordered* categories.

Section 2 proposes the measure to represent the degree of departure from the EBT model. Section 3 gives the approximate confidence interval for the measure. Section 4 shows examples.

2. Measure

Let for $i < j < k$,

$$G^{(1)}_{ijk} = \frac{G_{ijk}}{\sum_{s<t<u} G_{stu}}, \quad G^{(2)}_{ijk} = \frac{G_{kji}}{\sum_{s<t<u} G_{uts}}$$

where

$$\sum_{s<t<u} G_{stu} \neq 0, \quad \sum_{s<t<u} G_{uts} \neq 0, \quad G_{ijk} + G_{kji} \neq 0.$$

The EBT model may be expressed as

$$G^{(1)}_{ijk} = G^{(2)}_{ijk} \quad \text{for} \quad i < j < k.$$

Denote any probabilities having the structure of EBT by $\{q_{ij}\}$ with $q_{ij} + q_{ji} = 1$. Then denote $\{G^{(t)}_{ijk}\}$ with $\{\pi_{ij}\}$ replaced by $\{q_{ij}\}$, by $\{Q^{(t)}_{ijk}\}$, $t = 1, 2$. Thus

$$Q^{(1)}_{ijk} = Q^{(2)}_{ijk} \ (= Q^{EBT}_{ijk}) \quad \text{for} \quad i < j < k.$$

Consider a measure defined by

$$\Psi = \frac{1}{2\log 2} \min_{\{Q^{EBT}_{ijk}\}} \sum_{t=1}^{2} I\left(\left\{G^{(t)}_{ijk}\right\}; \left\{Q^{EBT}_{ijk}\right\}\right), \tag{1}$$

where

$$I\left(\left\{a_{ijk}\right\}; \left\{b_{ijk}\right\}\right) = \sum_{i<j<k} a_{ijk} \log\left(\frac{a_{ijk}}{b_{ijk}}\right),$$

being the Kullback-Leibler information. Then we can see that Q^{EBT}_{ijk} satisfying (1) are $\bar{Q}^{EBT}_{ijk} = (G^{(1)}_{ijk} + G^{(2)}_{ijk})/2$ for $i < j < k$. Thus, the measure can be expressed as

$$\Psi = \frac{1}{2\log 2} \sum_{t=1}^{2} I\left(\left\{G^{(t)}_{ijk}\right\}; \left\{\frac{G^{(1)}_{ijk} + G^{(2)}_{ijk}}{2}\right\}\right).$$

We see that (i) $0 \leq \Psi \leq 1$, (ii) $\Psi = 0$ if and only if the EBT model holds, and (iii) $\Psi = 1$ if and only if the degree of departure from the EBT model is maximum, in the sense that $G^{(1)}_{ijk} = 0$ (then $G^{(2)}_{ijk} > 0$) for some $i < j < k$ and $G^{(2)}_{ijk} = 0$ (then $G^{(1)}_{ijk} > 0$) for the other $i < j < k$. The maximum degree of departure from EBT can also be expressed as $G_{ijk}/(G_{ijk} + G_{kji}) = 0$ for some $i < j < k$ and $G_{kji}/(G_{ijk} + G_{kji}) = 0$ for the other $i < j < k$. Namely, $\Psi = 1$ indicates that for any three teams of R teams, the conditional probability that team i defeats team j, team j defeats team k, and team k defeats team i on conditional that i defeats j, j defeats k and k defeats i, or i defeats k, k defeats j, j defeats i, is 0 or 1. We shall refer to this situation as "strongest stochastic three way deadlock". Note that from the assumption, $G^{(1)}_{ijk} = 0$ for all $i < j < k$ and $G^{(2)}_{ijk} = 0$ for all $i < j < k$ are excluded from the strongest stochastic three way deadlock. Moreover, since $\Psi = 1$ indicates that $G^{(1)}_{ijk}/G^{(2)}_{ijk} = 0$ for some $i < j < k$ and $G^{(1)}_{ijk}/G^{(2)}_{ijk} = \infty$ for the other $i < j < k$, it seems appropriate to consider that then the degree of departure from EBT (i.e., from $G^{(1)}_{ijk}/G^{(2)}_{ijk} = 1$ for $i < j < k$) is the largest.

3. Approximate Confidence Interval for Measure

Consider a set of data from $R(R-1)/2$ paired comparison experiments for R treatments. Let r_{ij} be the number of comparisons for the treatment pair (i, j), and n_{ij} the number that the treatment i exceeds the treatment j in the r_{ij} comparisons. We assume that there is no tie, i.e., $r_{ij} = r_{ji} = n_{ij} + n_{ji}$. The probability for $\{n_{ij}\}$, $i \neq j$, is then the product of $R(R-1)/2$ binomials. The sample version of Ψ, i.e., $\hat{\Psi}$, is given by Ψ with $\{\pi_{ij}\}$ replaced by $\{\hat{\pi}_{ij}\}$, where $\hat{\pi}_{ij} = n_{ij}/r_{ij}$. Using the delta method (Bishop, Fienberg and Holland, 1975, sec.14.6), $\hat{\Psi}$ has asymptotically a normal distribution with mean Ψ and variance $\sigma^2[\hat{\Psi}]$. The $\sigma^2[\hat{\Psi}]$ is given in Appendix 1.

Let $\hat{\sigma}^2[\hat{\Psi}]$ denote $\sigma^2[\hat{\Psi}]$ with $\{\pi_{ij}\}$ replaced by $\{\hat{\pi}_{ij}\}$. Then, $\hat{\sigma}[\hat{\Psi}]/\sqrt{n}$ is an estimated approximate standard error for $\hat{\Psi}$, and $\hat{\Psi} \pm z_{p/2}\hat{\sigma}[\hat{\Psi}]/\sqrt{n}$ is an approximate $100(1-p)$ percent confidence interval for Ψ, where $z_{p/2}$ is the percentage point from the standard normal distribution corresponding to a two-tail probability equal to p.

4. Examples

Table 1 gives the results of professional baseball league in Japan in 2008 and 2011. These data are obtained from the official website of Japan Professional Baseball (http://www.npb.or.jp/). The categories have the ranking in these years. Namely, for the data in Table 1a, the first is Giants, the second is Tigers and so on. For example, from Giants's perspective, the (Giants, Tigers) result in 2008 correspond to 14 successes and 10 failures in 24 trials.

<Table 1>

The estimated measure $\hat{\Psi}$ are 0.137 for the data in Table 1a and 0.081 for the data in Table 1b. The approximate 95% confidence interval for Ψ are $(0.014, 0.259)$ with standard error 0.063 for the data in Table 1a and $(-0.021, 0.184)$ with standard error 0.052 for the data in Table 1b. Since the confidence interval for Ψ applied to the data in Table 1a do not contain zero, this would indicate that there is not a structure of EBT between the teams in Central league in 2008. On the other hand, since the confidence interval for the measure applied to the data in Table 1b contains zero, this would indicate that there is a structure of EBT between the teams in Central league in 2011; or if this is not the case, then it indicates that the degree of departure from the EBT model is slight.

When the degrees of departure from the EBT model in Tables 1a and 1b are compared using the estimated measures $\hat{\Psi}$, it is greater for Table 1a than for Table 1b. Namely, the data in Table 1a rather than in Table 1b is estimated to be close to the *maximum* departure from the EBT model.

5. Discussions

Consider an $R \times R$ square contingency table with same ordinal row and column classifications. Let p_{ij} denote the probability that an observation will fall in the ith row and the jth column of the table ($i = 1, \dots, R; j = 1, \dots, R$). Tomizawa (1984) proposed the extended quasi-symmetry (EQS) model defined by

$$p_{ij} = \alpha_i \beta_j \psi_{ij} \quad \text{for} \quad i = 1, \dots, R; j = 1, \dots, R,$$

where $\psi_{ij} = \gamma \psi_{ji}$ $(i < j)$. Let $p^c_{ij} = p_{ij}/(p_{ij} + p_{ji})$ for $i \neq j$. Then the EQS model may also be expressed as

$$p^c_{ij} p^c_{jk} p^c_{ki} = \gamma p^c_{ji} p^c_{kj} p^c_{ik} \quad \text{for} \quad i < j < k.$$

It is seen that the EQS model is essentially equivalent to the EBT model. Thus we shall define the measure ϕ which represents the degree of departure from the EQS model, by Ψ with $\{\pi_{ij}\}$ replaced by $\{p^c_{ij}\}$.

Let x_{ij} denote the observed frequency in the ith row and the jth column of the table ($i = 1, \dots, R; j = 1, \dots, R$). We assume that $\{x_{ij}\}$ have a multinomial distribution. Let $\hat{\phi}$ denote ϕ with $\{p_{ij}\}$ replaced by $\{\hat{p}_{ij}\}$ where $\hat{p}_{ij} = x_{ij}/n$ with $n = \sum\sum x_{ij}$. Using delta method, $\hat{\phi}$ has asymptotically a normal distribution with mean ϕ and variance $\sigma^2[\hat{\phi}]$. The measure $\hat{\Psi}$ is applied to the data obtained from independent binomial sampling, and $\hat{\phi}$ is applied to the data obtained from multinomial sampling. So, $\sigma^2[\hat{\Psi}]$ with $\{\pi_{ij}\}$ replaced by $\{p^c_{ij}\}$, $i \neq j$, is not always identical to $\sigma^2[\hat{\phi}]$. Let $\hat{\sigma}^2[\hat{\phi}]$ denote $\sigma^2[\hat{\phi}]$ with $\{p_{ij}\}$ replaced by $\{\hat{p}_{ij}\}$. Noting that $\{\hat{p}_{ij} + \hat{p}_{ji} = (x_{ij} + x_{ji})/n\}$ in $\hat{\sigma}^2[\hat{\phi}]$, we point out that the estimated variance $\hat{\sigma}^2[\hat{\phi}]$ is theoretically identical to the estimated variance $\hat{\sigma}^2[\hat{\Psi}]$. For more detail, see Tahata et al. (2004).

Note that we can consider a generalized measure for representing the degree of departure from the EBT (EQS) model by using the power-divergence (Cressie and Read, 1984) including the Kullback-Leibler information as follows: for $\lambda > -1$,

$$\Psi^{(\lambda)} = \frac{\lambda(\lambda+1)}{2(2^\lambda - 1)} \sum_{t=1}^{2} I^{(\lambda)}\left(\left\{G^{(t)}_{ijk}\right\}; \left\{\frac{G^{(1)}_{ijk} + G^{(2)}_{ijk}}{2}\right\}\right),$$

where

$$I^{(\lambda)}\left(\left\{a_{ijk}\right\}; \left\{b_{ijk}\right\}\right) = \frac{1}{\lambda(\lambda+1)} \sum_{i<j<k} a_{ijk} \left[\left(\frac{a_{ijk}}{b_{ijk}}\right)^\lambda - 1\right],$$

and the value at $\lambda = 0$ is taken to be the limit as $\lambda \to 0$. When $\lambda = 0$, $\Psi^{(0)}$ is identical to Ψ. The approximate variance of estimated measure $\hat{\Psi}^{(\lambda)}$ is given in Appendix 2.

Consider the data in Table 1, again. Since $\hat{\Psi} = 0.137$ for Table 1a, we can see that the degree of departure from EBT is estimated to be 13.7 percent of the maximum degree of departure from EBT. Similarly, we can infer that the degree of departure from EBT is 8.1 percent of the maximum degree of departure from EBT for the data in Table 1b. Also, we point out that the measure proposed in this paper may be useful to analyze the square contingency tables, for example, social mobility data, paired comparison data, and so on.

6. Concluding Remarks

Since the measure Ψ always ranges between 0 and 1 independent of the number of categories and sample size, it may be useful for comparing the degree of departure from the EBT model in several tables.

The proposed measures would be useful when we want to see with single summary measure what degree the departure from EBT is toward the strongest stochastic three way deadlock, although we cannot see it by the test statistic.

The proposed measures are not invariant under the arbitrary similar permutations of row and column categories. Therefore it is possible to apply these measures for analyzing the data on an *ordered* categories.

Finally, for the data having nominal category, if one wants to measure the degree of departure from **BT**, it is appropriate to use the measure proposed by Tahata et al. (2004). On the other hand, for the data having ordinal categories, if one wants to measure the degree of departure from **EBT**, it is appropriate to use the measure proposed.

Acknowledgements

The authors are grateful to the editor and a referee for their helpful comments and suggestions.

Appendix 1

The variance $\sigma^2[\hat{\Psi}]$ is given as follows:

$$\sigma^2[\hat{\Psi}] = \sum_{a=1}^{R-1} \sum_{b=a+1}^{R} \frac{1}{r_{ab}} \left\{ \frac{1}{\pi_{ab}} (A_{ab})^2 + \frac{1}{\pi_{ba}} (B_{ab})^2 - (A_{ab} + B_{ab})^2 \right\}, \tag{A.1}$$

where

$$\begin{aligned} A_{ab} &= \frac{1}{2\log 2} \sum_{i<j<k} \Big[G^{(1)}_{ijk} (\log H^{(1)}_{ijk}) \Big\{ I_{ij} + I_{jk} - \sum_{s<t<u} G^{(1)}_{stu} (I_{st} + I_{tu}) \Big\} \\ &\quad + G^{(2)}_{ijk} (\log H^{(2)}_{ijk}) \{ I_{ik} - \sum_{s<t<u} G^{(2)}_{stu} I_{su} \} \Big], \end{aligned}$$

with

$$H^{(t)}_{ijk} = \frac{G^{(t)}_{ijk}}{G^{(1)}_{ijk} + G^{(2)}_{ijk}}, \quad I_{ij} = \begin{cases} 1 & (\text{when} \quad i = a \quad \text{and} \quad j = b), \\ 0 & (\text{otherwise}), \end{cases}$$

and B_{ab} is defined by A_{ab} obtained by interchanging $G^{(1)}_{ijk}$ and $G^{(2)}_{ijk}$.

Appendix 2

The variance $\sigma^2[\hat{\Psi}^{(\lambda)}]$ is given by (A.1), where for $\lambda > -1$ and $\lambda \neq 0$,

$$\begin{aligned} A_{ab} &= \frac{2^{\lambda-1}}{2^{\lambda} - 1} \sum_{i<j<k} \Big[G^{(1)}_{ijk} (H^{(1)}_{ijk})^{\lambda} \Big\{ I_{ij} + I_{jk} - \sum_{s<t<u} G^{(1)}_{stu} (I_{st} + I_{tu}) \Big\} + G^{(2)}_{ijk} (H^{(2)}_{ijk})^{\lambda} \Big\{ I_{ik} - \sum_{s<t<u} G^{(2)}_{stu} I_{su} \Big\} \\ &\quad + \lambda \Big((H^{(1)}_{ijk})^{\lambda+1} G^{(2)}_{ijk} - (H^{(2)}_{ijk})^{\lambda+1} G^{(1)}_{ijk} \Big) \Big\{ I_{ij} + I_{jk} - I_{ik} - \sum_{s<t<u} \Big(G^{(1)}_{stu} I_{st} + G^{(1)}_{stu} I_{tu} - G^{(2)}_{stu} I_{su} \Big) \Big\} \Big], \end{aligned}$$

and B_{ab} is defined by A_{ab} obtained by interchanging $G^{(1)}_{ijk}$ and $G^{(2)}_{ijk}$. When $\lambda = 0$, $\sigma^2[\hat{\Psi}^{(0)}]$ is identical to $\sigma^2[\hat{\Psi}]$.

References

Agresti, A. (1990). *Categorical Data Analysis.* New York: Wiley.

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

Bradley, R. A., & Terry, M. E. (1952). Rank analysis of incomplete block designs I. The method of paired comparisons. *Biometrika, 39*, 324-345.

Cressie, N. A. C., & Read, T. R. C. (1984). Multinomial goodness-of-fit tests. *Journal of the Royal Statistical Society, Series B, 46*, 440-464.

Davidson, R. R., & Beaver, R. J. (1977). On extending the Bradley-Terry model to incorporate within-pair order effects. *Biometrics, 33*, 693-702. http://dx.doi.org/10.2307/2529467

Tahata, K., Miyamoto, N., & Tomizawa, S. (2004). Measure of departure from quasi-symmetry and Bradley-Terry models for square contingency tables with nominal categories. *Journal of the Korean Statistical Society, 33*, 129-147.

Tomizawa, S. (1984). Three kinds of decompositions for the conditional symmetry model in a square contingency table. *Journal of the Japan Statistical Society, 14*, 35-42.

Table 1. Score sheet of the Central League in Japan in 2008 and 2011

(a) 2008

	Giants	*Tigers*	*Dragons*	*Carp*	*Swallows*	*Baystars*	*Total*
Giants	-	14	10	10	18	18	70
Tigers	10	-	17	14	13	13	67
Dragons	14	6	-	13	9	17	59
Carp	12	10	9	-	12	13	56
Swallows	6	10	13	11	-	15	55
Baystars	5	10	7	11	9	-	42
Total	47	50	56	59	61	76	349

(b) 2011

	Dragons	*Swallows*	*Giants*	*Tigers*	*Carp*	*Baystars*	*Total*
Dragons	-	11	10	13	12	15	61
Swallows	10	-	12	10	13	15	60
Giants	12	8	-	11	16	14	61
Tigers	9	14	11	-	12	12	58
Carp	10	9	6	12	-	17	54
Baystars	8	5	10	10	7	-	40
Total	49	47	49	56	60	73	334