International Journal of Probability and Statistics 2013, 2(3): 43-49
DOI: 10.5923/j.ijps.20130203.01

Bayesian Inference of Log-linear Version of the Bradley-Terry Model for Paired Comparisons Using Uninformative Prior

Syed Hussain[1,*], **Muhammad Aslam**[2]

[1]Master's Degree, Lecturer in University of Gujrat, Gujrat, Pakistan
[2]Doctor of Philosophy, Professor in Quaid-i-Azam University Islamabad, Pakistan

Abstract Bayesian analysis of log-linear version of the Bradley-Terry[3] model is performed in this paper considering generalization of Dittrich et al.,[6]; Dittrich et al.,[7] and the Dittrich et al.,[8] to modify and re-estimate the model parameters to overcome a small deficiency in the estimation of a single log odd parameter being aliased. To ensure ranking is maintained, we computed the posterior predictive probabilities and posterior probabilities of hypotheses as per the criteria by Aslam, M[2].

Keywords Method of Paired Comparisons, Bayesian Statistics

1. Introduction

The method of paired comparisons provides us basis for comparing the objects or stimuli in the form of pairs to obtain ranks. A detailed discussion on the method is given in David[5]. In paired comparison experiments, a judge or a panel of judges examine pairs of objects. The worth or merit of an object is measured through comparisons against other units. Thurstone[15] presented a major advance in the field of Psychometric scaling; a science that determines measuring techniques for human judgments. In this perspective, Bradley-Terry[3] published an alternative version of the paired comparison model developed by Thurstone[15]. Under the Bradley-Terry model for two objects with worth parameters θ_i and θ_j, the preferences probability for the object, A_i and A_j are below as;

$$\pi_{ij} = \frac{\theta_i}{\theta_i + \theta_j}$$

$$\pi_{ji} = \frac{\theta_j}{\theta_i + \theta_j}$$

Alternatively the Bradley-Terry Model can be fitted as log linear model (see e.g., 1; 6; 10; 14). Dittrich &Hatzinger[8] fitted the log-linear version of Bradley-Terry[3] Model using R package (see 7) after the formulations of the following equations:

$$m_{ij} = n_{ij}.\pi_{ij} = n_{ij}.\frac{\sqrt{\frac{\theta_i}{\theta_j}}}{\sqrt{\frac{\theta_i}{\theta_j}} + \sqrt{\frac{\theta_j}{\theta_i}}}$$

$$\ln(m_{ij}) = \mu_{ij} + \gamma_i + \gamma_j$$

$$\ln(m_{ji}) = \mu_{ji} + \gamma_i + \gamma_j$$

Where the nuisance parameter μ_{ij} may be interpreted as interaction parameters representing the universities involved in the respective comparisons and the universities related terms are denoted by γ_i and γ_j.

2. Modification of Log-Linear Bradley-Terry Model

The basic Bradley-Terry model is invariant under the change of scale and identification is obtained under the condition:

$$\sum_{j=1}^{m} \theta_j$$

Using $\gamma_j = \frac{1}{2}\ln(\theta_j)$ and $\theta_j = e^{2\gamma_j}$ we get,

$$\theta_j = \frac{e^{2\gamma_j}}{1} \Rightarrow \theta_j = e^{2\gamma_j} / \sum_{j=1}^{m} e^{2\gamma_j}$$

---

* Corresponding author:
syed.adil@uog.edu.pk (Syed Hussain)
Published online at http://journal.sapub.org/ijps

After simplifying; we obtain some modified form, after the log-linear Bradley-Terry model by Dittrich et al.[6], for the estimation of log odds via Bayesian paradigm as below:

$$\psi_{ij} = \frac{e^{2\gamma_i}}{e^{2\gamma_i} + e^{2\gamma_j}}$$

$$\psi_{ji} = \frac{e^{2\gamma_j}}{e^{2\gamma_i} + e^{2\gamma_j}}$$

The ψ_{ij} and ψ_{ji} are the preferences probabilities based on the log odds parameters.

3. Bayesian Inference of the Modified Function

Bayesian analyses of log-linear models are complicated as we usually perform these analyses using complicated numerical integrations. It is of great practice that the posterior distribution turns out to be an improper density function. Therefore, we consider the non-informative Jeffreys' prior (see 11 & 4) for the proposed model. This analysis is based on the likelihood function and the prior distribution; we first derive the likelihood and define the prior distribution as below.

3.1. Notations and Likelihood Function

In the present situation, we see that there are only two possible outcomes of the paired comparison experiment, i.e., either object, 'A_i' is preferred to 'A_j' or the vice versa. The preference probability ψ_{ij} denotes the probability of object, 'A_i' preferred over object 'A_j' in all 'z_{ij}' fixed number of independent paired comparisons for all of the pair of objects. Random variable $X_{i.ij}$ is assumed to follow a binomial distribution $B\left(x_{i.ij}; z_{ij}, \psi_{ij}\right)$ and the likelihood function takes the form:

$$L(X;\ \gamma_1, \gamma_2, \ldots, \gamma_m) = \frac{\prod_{i=1}^{m}\left(e^{2\gamma_i}\right)^{n_i}}{k\prod_{i<j}^{m}\left(e^{2\gamma_i} + e^{2\gamma_j}\right)^{z_{ij}}} \quad \gamma_6 = 1 - \gamma_1 - \gamma_2 - \gamma_3 - \gamma_4 - \gamma_5$$

The 'γ_6' is the constraint on the numerical integration and 'k' is the normalizing constant and 'n_i' is the total number of times object 'A_i' is preferred.

3.2. Jeffreys' Prior Distribution

The Jeffreys' prior[11] for the '$\underline{\gamma}$' is proportional to the square root of determinant of Fisher (9) Information Matrix and given as:

$$\underset{J}{P}\left(\underline{\gamma}\right) = \sqrt{\det\left[I_{p.p}\left(\underline{\gamma}\right)\right]}$$

The $\underline{\gamma} = \gamma_1, \gamma_2, \ldots, \gamma_m$

The $I_{p.p}\left(\underline{\gamma}\right)$ be the 'p x p' Fisher[9] information matrix, that is, the logarithm of likelihood functions of parameter space $\underline{\gamma}$ and given as follow:

$$I_{pq}\left(\underline{\gamma}\right) = -E\left\{\frac{\partial^2 \log L\left(\underline{\gamma}\right)}{\partial\gamma_p \partial\gamma_q}\right\}$$

The $L\left(\underline{\gamma}\right)$ represents the likelihood function.

3.3. Posterior Distribution for the Model via the Jeffreys Prior

The joint posterior distribution with the Jeffreys'[11] prior takes the following form

$$\mathrm{P}(\underline{\gamma}|x) = \frac{\prod_{i=1}^{6}\left(e^{2\gamma_i}\right)^{n_i} \underset{J}{P}\left(\underline{\gamma}\right)}{K\left(\prod_{i<j}^{6}\left(e^{2\gamma_i} + e^{2\gamma_j}\right)^{z_{ij}}\right)}$$

The ‘ K ’ is the normalizing constant. The identifiably condition is obtained with:

$$\sum_{i=1}^{m} \gamma_i = 1$$

We need the following second order partial derivatives of Log-likelihood functions of paired comparison model.

$$f_{uu}^{//} = \frac{\partial^2 f(u,v)}{\partial u^2}\Big|_{(u_o,v_o)} = \left[\frac{f(u_o - k, v_o) - 2f(u_o, v_o) + f(u_o + k, v_o)}{k^2} \right]$$

$$f_{uv}^{//} = \left[\frac{f(u_o + k, v_o + l) - 2f(u_o = k, v_o = l) - k^2 (f_{uu}^{//} + f_{uv}^{//}) - 2f(u_o, v_o)}{2k^2} \right]$$

In Table 1, data is taken from Dittrich et al.[6] about students’ preferences for the six European universities as:

Table 1. Students’ Preferences Data of European Universities

Pairs	$x_{i.ij}$	$x_{j.ij}$	z_{ij}
(1,2)	34	16	50
(1,3)	40	10	50
(1,4)	37	13	50
(1,5)	38	12	50
(1,6)	44	6	50
(2,3)	34	16	50
(2,4)	29	21	50
(2,5)	30	20	50
(2,6)	37	13	50
(3,4)	24	26	50
(3,5)	22	28	50
(3,6)	31	19	50
(4,5)	26	24	50
(4,6)	31	19	50
(5,6)	33	17	50

Source: Dittrich et al. (1998)

3.4. Bayesian Estimation Using Modified Form of the Model Parameters

Table 2. Baye’s Estimators of Modified model

Parameter	Posterior Means	Estimate by Dittrich et al. (1998)
γ_1 (LO)	0.57346	0.7906
γ_2 (PA)	0.20719	0.3974
γ_3 (MI)	0.037468	0.1045
γ_4 (SG)	0.082754	0.1820
γ_5 (BA)	0.075303	0.0805
γ_6 (ST)	0.05237	0.0 (Aliased)

Table 2 shows estimates for the Posterior Means and MLE (see 6). The log odds differs only in account of one aliased parameter by Dittrich et al.[6] while as Bayesian analysis generates that one using Jeffreys' prior.

3.5. Posterior Predictive Probabilities Using Jeffrey's Prior

The predictive probabilities[13] show the preferences for the pair of objects when a single comparison between pair of objects will carry out in the future. The formula for predictive probability for objects A_i and A_j is given as below:

$$\mathrm{P}(ij) = \frac{1}{C}\int_{\gamma_1=0}^{1}\int_{\gamma_2=0}^{1-\gamma_1}\int_{\gamma_3=0}^{1-\gamma_1-\gamma_2}\int_{\gamma_4=0}^{1-\gamma_1-\gamma_2-\gamma_3}\int_{\gamma_5=0}^{1-\gamma_1-\gamma_2-\gamma_3-\gamma_4} \mathrm{P}(\underline{\gamma}|\underline{x}).\psi_{ij}\,d\gamma_5 d\gamma_4 d\gamma_3 d\gamma_2 d\gamma_1$$

$$\mathrm{P}(\underline{\gamma}|\underline{x}) = \left[\frac{\prod_{i=1}^{6}(e^{2\gamma_i})^{n_i}}{k\prod_{i<j}^{6}(e^{2\gamma_i}+e^{2\gamma_j})^{z_{ij}}}.P_J(\underline{\gamma})\right]$$

$$\psi_{ij} = \left(\frac{e^{2\gamma i}}{e^{2\gamma_i}+e^{2\gamma_j}}\right)$$

Table 3 shows the posterior predictive probabilities for fifteen pair of objects it also indicates the following relationship between the six universities at the scale of preferences.

$$\gamma_1 \to \gamma_2 \to \gamma_6 \to \gamma_5 \to \gamma_4 \to \gamma_3$$

Table 3. Estimates of Posterior Predictive Probabilities

Predictive Probabilities Parameters	$\mathrm{P}(ij)$	$\mathrm{P}(ji)$
$\mathrm{P}(12)$	0.67424	0.32576
$\mathrm{P}(13)$	0.74431	0.25569
$\mathrm{P}(14)$	0.7263	0.2737
$\mathrm{P}(15)$	0.72928	0.27072
$\mathrm{P}(16)$	0.74959	0.25041
$\mathrm{P}(23)$	0.58379	0.41621
$\mathrm{P}(24)$	0.56156	0.43844
$\mathrm{P}(25)$	0.56522	0.43478
$\mathrm{P}(26)$	0.50679	0.49321
$\mathrm{P}(34)$	0.47742	0.52258
$\mathrm{P}(35)$	0.48113	0.51887
$\mathrm{P}(36)$	0.50679	0.49321
$\mathrm{P}(45)$	0.50369	0.49631
$\mathrm{P}(46)$	0.52933	0.47067
$\mathrm{P}(56)$	0.52563	0.47437

3.6. Bayesian Testing of Hypothesis Using Jeffrey's Prior

The null and alternate hypotheses for pair of objects are

$$H_{ij};\ \gamma_i > \gamma_j$$

$$\overline{H}_{ij};\ \gamma_j \geq \gamma_i \quad i<j=1,\ldots,6$$

The general formula to calculate the posterior probabilities of null hypothesis for objects A_i with A_j is given below as:

$$P(\phi > 0|\mathrm{x}) = p_{ij} = \int_{\phi=0}^{1}\int_{\zeta=\phi}^{(1+\phi)/2}\int_{\gamma_k=0}^{1-2\zeta+\phi}\int_{\gamma_l=0}^{1-2\zeta+\phi-\gamma_k}\int_{\gamma_m=1}^{1-2\zeta+\phi-\gamma_k-\gamma_l} P\left(\phi,\zeta,\gamma_k,\gamma_l,\gamma_m|\mathrm{x}\right)d\gamma_m d\gamma_l d\gamma_k d\zeta d\phi$$

With $q_{ij} = 1 - p_{ij}$ and $P_J\left(\underline{\gamma}\right) \propto \sqrt{I\left(\underline{\gamma}\right)}$

And $I_{rc}\left(\underline{\gamma}\right) = -E\left\{\dfrac{\partial^2 \log L\left(\underline{\gamma};x\right)}{\partial\gamma_r \partial\gamma_c}\right\}$

Here $P\left(\phi,\zeta,\gamma_k,\gamma_l,\gamma_m|\mathrm{x}\right) \propto L\left(\underline{\gamma};x\right) \; P_J\left(\underline{\gamma}\right)$ $k,l,m \neq i.j$

And $\gamma_n = 1 - 2.0*\phi + \zeta - \gamma_k - \gamma_l - \gamma_m$

We use the following transformation by Aslam[2] to obtain the posterior probabilities of the hypotheses.

$$\gamma_i = \phi$$

$$\gamma_j = \phi - \zeta$$

We follow the decision criteria suggested by Aslam[2]. The criteria are easy to understand that is if any one of the posterior probability of hypothesis either p_{ij} of H_{ij} or q_{ij} of $\overline{H}_{ij}$ is more than 90%, that hypothesis will be accepted. The posterior probabilities of hypotheses are shown in Table 4. We denote the posterior probabilities by p_{ij} in favor of i^{th} objects and q_{ij} in favor of j^{th} objects. We now interpret each of the tested hypotheses for six parameters. The probability p_{12} of H_{12} shows the strongest favor of London University when compared with Paris University. Also q_{12} has the probability less than 10 %, so we accept the hypothesis H_{12} and conclude that London University has the greater preference probability when compared with Paris University. Now q_{13} has the probability, which is less than 10% so we accept H_{13} and we conclude that London University has the greater preference probability when compared with Milan University. Also the decision for the greater preference probability of London School of Economics and St. Gallen University is inconclusive. Also the decisions for the greater preference probabilities of London School of Economics against Barcelona University and Stockholm School of Economics are inconclusive. The probability for $\overline{H}_{23}$ is less than 10%, so we accept H_{23} and conclude that Paris University has the greater preference probability when compared with Milan University. The probabilities of $\overline{H}_{24}$, $\overline{H}_{25}$ and $\overline{H}_{26}$ are less than 10%, so we accept H_{24}, H_{25} and H_{26} concluding that Paris University has the greater preference probabilities when compared with St. Gallen University, Barcelona University and Stockholm School of Economics. The probability for H_{34} is not less than 10%, so we could not accept any of the hypotheses and the decision is inconclusive. The probability for H_{35} is also greater than 10%, so decision is inconclusive. The probability for H_{36} is less than 10 %, so we accept $\overline{H}_{36}$ and conclude that Stockholm School of Economics has the greater preference probability when compared with Milan University. The probability for H_{45} is not less than 10% so the decision is inconclusive. The probability for H_{46} is not less than 10 %, so the decision is inconclusive. The probability for H_{56} is less than 10 %, so we accept $\overline{H}_{56}$ and conclude that Stockholm School of Economics has the greater preference probability when compared with Barcelona University.

Table 4. Estimates of Posterior Probabilities of Hypothesis

Pairs	p_{ij}	q_{ij}
(1,2)	1	0
(1,3)	1	0
(1,4)	1	0
(1,5)	1	0
(1,6)	1	0
(2,3)	0.99987	0.00012
(2,4)	0.99996	0.00003
(2,5)	0.99998	0.00001
(2,6)	0.99783	0.00216
(3,4)	0.14289	0.85710
(3,5)	0.59727	0.40272
(3,6)	0.06361	0.93638
(4,5)	0.84372	0.15627
(4,6)	0.35910	0.64089
(5,6)	0.000098	0.99999

These probabilities of hypothesis ensure us that the estimates are correct. Among the p_{ij}'s, six null and alternative hypotheses are accepted having the strong probability of acceptance, while as, three hypotheses are remained inconclusive.

3.7. Appropriateness of the Model

The classical technique of Chi-Square method to test the hypothesis of goodness of fit for the modified form of the model is used. The null and alternate hypotheses are as follow:

H_o; The model is good fit of the data

$\overline{H}_o$; The model does not fit the data

We calculate the expected frequencies by the following formula:

$$\hat{x}_{ij} = z_{ij}(\psi_{ij}) \quad i < j$$

The level of significance is 5% and the test statistic follows the Chi-Square distribution as:

$$\chi^2 = \sum_{i<j=1}^{m} \left\{ \frac{(x_{ij} - \hat{x}_{ij})^2}{\hat{x}_{ij}} + \frac{(x_{ji} - \hat{x}_{ji})^2}{\hat{x}_{ji}} \right\}$$

We follow the consideration by Aslam[2] for the choice of degree of freedom by the following formula:

$$d.f = m(m-2) = 24.$$

Table 5 shows the observed and expected number of preferences as below in Table 5 as follow:

Table 5. Estimates of Observed and Expected Frequencies

x_{ij}	$\hat{x}_{ij}$	x_{ji}	$\hat{x}_{ji}$
34	33.3575	16	16.6426
40	37.5285	10	12.4715
37	36.9913	13	13.0088
38	36.9913	12	13.0088
44	37.0555	6	12.9446
34	30.0108	16	19.9893
29	29.3275	21	20.6725
30	29.3275	20	20.6725
37	29.4085	13	20.5915
24	22925	26	25.7076
22	22925	28	25.7076
31	23759	19	25.6241
26	25	24	25
31	25.0835	19	29165
33	25.0835	17	29165

In Table 5, Chi-Square test statistic is computed as:

$$\chi^2_{cal} = 23.78784331.$$

With p-value= 0.473781549.
And the table value is:

$$\chi^2_{(0.05,24)} = 36.42.$$

Critical Region is as follow:

$$\chi^2_{cal} \not\geq \chi^2_{(0.05,24)}$$

From the critical region, there is no evidence to reject the null hypothesis; therefore, we conclude that the model good fits the data.

4. Conclusions & Discussions

Bayesian inference using Jeffreys[11] prior produced consistent estimates as compare to the classical approach by overcoming the little deficiency in the estimation of a parameter aliased by the estimation technique of Dittrich et al.[6] (see Table 2). Posterior predictive probabilities for each pair of Universities are obtained in Table 3. Ranking is ensured in Table 2 through posterior probabilities of hypotheses for each pair of Universities in Table 4.

Posterior means for object related parameters $\underline{\gamma}$ are obtained for ranking the six European Universities. It could be further generalized for the parameters of ties, order effects, the object specific covariates, subject specific covariates and their interaction parameters via Bayesian inference.

REFERENCES

[1] Agresti, A. (1990). Categorical Data Analysis. J.Wiley, New York.

[2] Aslam, M. (2002). "Bayesian Analysis for the paired Comparison Models allowing Ties and not allowing Ties". Pakistan Journal of Statistics Vol. 18, No. 1, pp. 53-69.

[3] Bradley, R. A. and Terry, M. (1952). "Rank Analysis of Incomplete Block Designs: The Method of Paired Comparisons". Biometrika Vol. 39(3-4): pp. 324–345.

[4] Berger, J. O. (1985). "Statistical Decision Theory and Bayesian Analysis". 2nd ed. Springer-Verlag: New York. 386.

[5] David, H.A. (1988). The Method of Paired Comparisons. Second ed. Charles Griffin & Company Ltd., London.

[6] Dittrich, R., Hatzinger, R. and Katzenbeisser, W. (1998). "Modeling the Effect of Subject-Specific Covariates in Paired Comparison Studies with an Application to University Rankings". Journal of the Royal Statistical Society Series C (Applied Statistics) Vol. 47, No. 4, pp. 511-525.

[7] Dittrich, R. & Hatzinger, R. (2009). "Fitting loglinear Bradley-Terry models (LLBT) for paired comparisons using the R package prefmod". Psychology Science Quarterly, Volume 51, (2), pp. 216 – 242.

[8] Dittrich, R. and Hatzinger, R. (2012). "prefmod: An R Package for Modeling Preferences based on Paired Comparisons, Rankings, or Ratings". Volume 48, Issue 10.

[9] Edgeworth, F. Y. (1908). "On the Probable Errors of Frequency-Constants" Journal of the Royal Statistical Society 71 (3): 499–512.

[10] Fienberg, S.E., & Larntz, K. (1976). "Loglinear representation for paired and multiple comparison models". Biometrika, 63, 245-254.

[11] Jeffreys, H. (1961). "Theory of Probability". Oxford University Press, London. 386, 387.

[12] Leonard, T. (1975). “Bayesian Estimation Methods for Two-Way Contingency Tables”. Journal of the Royal Statistical Society. Series B (Methodological) Vol. 37, No. 1, pp. 23–37.

[13] Lee, P. M. (1989). Bayesian Statistics: an Introduction. A Charles Griffin Book, Oxford University Press, New York.

[14] Sinclair, C.D. (1982). GLIM for preference. In: Gilchrist, R. (Eds.): GLIM 82. Proceedings of the International Conference on Generalized Linear Models, Springer Lecture Notes in Statistics, 14.

[15] Thurstone, L.L. (1927). A law of comparative judgment. Psychological Review, 34,273-286.