July 2012 IJPSS Volume 2, Issue 7 ISSN: 2249-5894

A BAYESIAN LOOK AT THE BRADLEY-TERRY PAIRED COMPARISON MODEL

Nosheen Altaf*

Ghausia Masood Gilani**

Muhammad Aslam***

ABSTRACT

In the present study, we have presented a statistical analysis for paired comparison data through Bayesian approach. In the method of paired comparisons, objects are presented in pairs to the judges. The Bradley-Terry model for paired comparisons is considered for Bayesian analysis. The prior distribution for the parameters of the model is supposed to be belonged to the member of the Dirichlet family and the method for elicitation of hyperparameters is based on the prior predictive distribution. For Bayesian analysis, the posterior distribution of the parameters is derived, the preference probabilities using the posterior means and the predictive probabilities for pairwise comparisons in a single future comparison are obtained. The posterior probabilities for the hypotheses of comparing two parameters are also computed. The goodness of fit test for the appropriateness of the model is presented too.

KEYWORDS: Paired comparison method; Bradley-Terry model; Dirichlet distribution; Hyperparameters; Elicitation; Posterior distribution; Bayesian hypothesis testing; Predictive probabilities; Preference probabilities.

* Riphah International University, Islamabad-44000, Pakistan.

** College of Statistical and Actuarial Sciences, University of the Punjab, Lahore, Pakistan.

*** Department of Statistics, Quaid-i-Azam University Islambad-44000, Pakistan.

1. INTRODUCTION

In the method of paired comparison, objects are presented in pairs to one or more judges for the purpose of comparison. The 'object' may be 'a person', 'a treatment', 'stimuli' and the like. The basic experimental unit is the comparison of two objects. Kim and Kim (2004) present a Bayesian approach to paired comparison of several products of Poisson Rates. Adams (2005) describes Bayesian versions of the method of paired comparison and their advantages of behavioral studies. Jian (2007) has developed statistical model and procedure for similarity tests using paired comparisons method. Miranda et al (2009) propose the use of a modified paired comparison method in which a reduced number of comparisons is selected according to an incomplete cyclic design. Causeu and Husson (2005) propose a 2-dimensional extension of the Bradley–Terry model. The detail about the paired comparisons method can be seen in David (1988).

The Bradley-Terry (1952) model is defined in Section 2 with the notations. Discussion about the informative prior and elicitation of hyperparameters are presented in Section 3. Bayesian analysis of the Bradley-Terry model for the number of treatments m=4 is given in section 4, this section includes formation of the posterior distribution, the marginal posterior densities, the posterior means and the modes of the parameters, the posterior probabilities of hypotheses for comparing two parameters, the predictive probabilities that one treatment would be preferred to another treatment in a future single comparison and the preference probabilities. Appropriateness of the model is also tested in Section 5.

2. THE BRADLEY-TERRY MODEL FOR PAIRED COMPARISON

The Bradley-Terry model introduced by Zermelo (1929) and developed by Bradley and Terry (1952). This model implies that the difference between two latent variables ($X_i - X_j$) has a logistic density with parameter ($\ln\theta_i - \ln\theta_j$). If ψ_{ij} denotes the probability $P(X_i > X_j \,|\theta_i,\theta_j)$ that the treatment T_i is preferred to the treatment T_j ($i \neq j$) when the treatments T_i and T_j are compared then

$$\psi_{ij} = \frac{1}{4}\int_{-(\ln\theta_i - \ln\theta_j)}^{\infty} \operatorname{sech}^2(y/2)dy \qquad i \neq j \text{, i,j=1,2,…,m.}$$

$$= \int_{-(\ln\theta_i - \ln\theta_j)}^{\infty} \frac{e^{-y}}{(1+e^{-y})^2}dy$$

$$= \frac{\theta_i}{\theta_i + \theta_j}. \qquad (2.1)$$

The Bradley-Terry model is defined by (2.1)

The following notations are used in the analysis:

n_{ijk} = 1 or 0 according as treatment T_i is preferred to treatment T_j or not in the k'th repetition (k=1,2, …, r_{ij}) of the comparison.

r_{ij} = the number of times treatment T_i is compared with treatment T_j.

$n_{ij} = \sum_k n_{ijk}$ = the number of times T_i is preferred to treatment T_j.

$n_i = \sum_{j\neq i}^{m} n_{ij}$ = the total number of times T_i is preferred to any other treatment.

Here $n_{ijk} + n_{jik} = 1$ and $r_{ij} = n_{ij} + n_{ji}$

Hence, the likelihood function of the observed outcome $\mathbf{x}$ which represents the data (r_{ij}, n_{ij}, n_{ji}) of the trial is:

$$l(\mathbf{x};\theta_1,...,\theta_m) = \prod_k \prod_{i<j}^{m} P_{ijk} = \prod_{i<j}^{m} \binom{r_{ij}}{n_{ij}} \frac{\theta_i^{n_{ij}}\theta_j^{n_{ji}}}{(\theta_i+\theta_j)^{r_{ij}}}$$

$$= \prod_{i=1}^{m} \theta_i^{n_{i.ij}} \prod_{i<j=1}^{m} \binom{r_{ij}}{n_{ij}} \theta_i + \theta_j^{\ -r_{ij}} . \qquad (2.2)$$

where θ_i *(i=1,2,…,m)* are the treatment parameters. We include a constraint on the parameters of the model that they are positive and sum to unity i.e. $\sum_{i=1}^{m}\theta_i = 1$, this condition ensures that parameters are well defined and identifiable.

3. 1 CHOICE OF AN INFORMATIVE PRIOR

Generally, the prior distribution is chosen with accord to the range of the parameter. Davidson and Solomon (1973) assume the natural conjugate family of prior for the parameters of the Bradley-Terry model. Leonard (1977) assumes the multivariate normal distribution as a prior distribution for the logarithm of the parameters. But neither discusses the method of elicitation of the hyperparameters.

In present study, prior distribution of the parameters: $\theta_1, \theta_2, ..., \theta_m$ is supposed to be belonged to the member of the Dirichlet family. i.e.

$$p(\boldsymbol{\theta}) = \frac{\Gamma(a_1 + ... + a_m)}{\Gamma(a_1)...\Gamma(a_m)} \prod_{i=1}^{m} \theta_i^{a_i - 1}, \quad \sum_{i=1}^{m} \theta_i = 1, \; a_i > 0; \quad i = 1, 2, ..., m \qquad (3.1)$$

where $\boldsymbol{\theta} = (\theta_1, \theta_2, ..., \theta_m)$ are the treatment parameters and $a_i \, (i = 1, 2, ..., m)$ are the hyperparameters.

3.2 Elicitation of Hyperparameters

Aslam (2003) describes three methods of eliciting hyperparameters via prior predictive distribution (PPD). The other authors who adopted the PPD for elicitation of prior distributions are Kadane et. al. (1980), Winkler (1980) and Chaloner and Duncan (1983). One of the methods to elicit the hyperparameters based on confidence levels of the PPD. The following function is used to elicit the hyperparameters:

$$\xi(a_1, ..., a_m) = \min_{a_1, ..., a_m} \sum_{k=1}^{l} \left| (CCL)_k - (ECL)_k \right| \qquad (3.2)$$

where k is the number of intervals considered in the elicitation, CCL is the confidence level characterized by the hyperparameters and ECL is the elicited confidence level. The set of hyperparameters with minimum value of $\xi(a_1, ..., a_m)$ in (3.2) is considered as the elicited values of hyperparameters.

For the number of times treatment T_i is preferred to treatment T_j when a pair of treatments (T_i, T_j) is being compared with the number of times 'r_{ij}', the prior predictive distribution { $p(n_{ij})$ } of the Bradley-Terry model for a pair (T_i, T_j) {using (2.1) and (3.1)} is:

$$p(n_{ij}) = K\int_0^1\int_0^{1-\theta_j} \frac{\theta_i^{n_{ij}+a_i-1}\theta_j^{n_{ji}+a_j-1}(1-\theta_i-\theta_j)^{c_k-1}}{(\theta_i+\theta_j)^{r_{ij}}} d\theta_i d\theta_j\,, \quad n_{ij} = 0,1...r_{ij} \qquad (3.3)$$

where $K = \frac{\Gamma(r_{ij}+1)\Gamma(a_i+a_j+c_k)}{\Gamma(n_{ij}+1)\Gamma(n_{ji}+1).\Gamma(a_i)\Gamma(a_j)\Gamma(c_k)}$, $n_{ji} = r_{ij} - n_{ij}$ and $c_k = \sum_{l\neq i,j}^{m} a_i$.

$$p(n_{ij}) = \frac{\Gamma(r_{ij}+1)B(a_i+n_{ij}, a_j+n_{ji})}{\Gamma(n_{ij}+1)\Gamma(n_{ji}+1).B(a_i,a_j)}\,, \qquad (3.4)$$

Here B stands for Beta function. [For more detail see Aslam (2003)].

For elicitation of hyperparameters, we assume a balanced design that the number of comparison for each pair is equal $(r_{ij} = 10)$; (i<j=1,2,3,4) and the expert confidence levels are given right side of the following equations. These equations are derived using the PPD (3.4):

$$\sum_{n_{12}=7}^{10} \frac{r_{12}!B(a_1+n_{12}, a_2+n_{21})}{n_{12}!(r_{12}-n_{12})!B(a_1,a_2)} = 0.15 \qquad (3.5)$$

$$\sum_{n_{13}=0}^{2} \frac{r_{13}!B(a_1+n_{13}, a_3+n_{31})}{n_{13}!(r_{13}-n_{13})!B(a_1,a_3)} = 0.15 \qquad (3.6)$$

$$\sum_{n_{23}=7}^{10} \frac{r_{23}!B(a_2+n_{23}, a_3+n_{32})}{n_{23}!(r_{23}-n_{23})!B(a_2,a_3)} = 0.40 \qquad (3.7)$$

$$\sum_{n_{14}=0}^{2} \frac{r_{14}!B(a_1+n_{14}, a_4+n_{41})}{n_{14}!(r_{14}-n_{14})!B(a_1,a_4)} = 0.70 \qquad (3.8)$$

$$\sum_{n_{24}=0}^{2} \frac{r_{24}!B(a_2+n_{24}, a_4+n_{42})}{n_{24}!(r_{24}-n_{24})!B(a_2,a_4)} = 0.15 \qquad (3.9)$$

$$\sum_{n_{34}=7}^{10} \frac{r_{34}!B(a_3+n_{34}, a_4+n_{43})}{n_{34}!(r_{34}-n_{34})!B(a_3,a_4)} = 0.65 \qquad (3.10)$$

A program is designed in SAS package to solve above equations to elicit the values of hyperparameters a_1, a_2, a_3 and a_4 which are found to be 0.4451, 0.8944, 0.7129 and 0.5567 respectively.

4.1 BAYESIAN ANALYSIS FOR THE BRADLEY-TERRY MODEL (*m=4*)

Let the number of treatments *m* be 4, here the treatment parameters are θ_1, θ_2, θ_3 and θ_4, using the constraint: $\theta_4 = 1-\theta_1-\theta_2-\theta_3$, now the likelihood is:

$$l(\mathbf{x};\, \theta_1,\theta_2,\theta_3) \propto \frac{\theta_1^{n_1}\theta_2^{n_2}\theta_3^{n_3}(1-\theta_1-\theta_2-\theta_3)^{n_4}}{(\theta_1+\theta_2)^{r_{12}}(\theta_2+\theta_3)^{r_{23}}(\theta_1+\theta_3)^{r_{13}}(1-\theta_2-\theta_3)^{r_{14}}(1-\theta_1-\theta_3)^{r_{24}}(1-\theta_1-\theta_2)^{r_{34}}} \quad (4.1)$$

where $n_1 = n_{1.12}+n_{1.13}+n_{1.14}$, $n_2 = n_{2.12}+n_{2.23}+n_{2.24}$, $n_3 = n_{3.13}+n_{3.23}+n_{3.34}$,

$n_4 = n_{4.14}+n_{4.24}+n_{4.34}$ and $r_{12} = n_{1.12}+n_{2.12}$, so on.

The prior distribution for *m=4* using the constraint: $\theta_4 = 1-\theta_1-\theta_2-\theta_3$, is:

$$p(\boldsymbol{\theta}) = \frac{\Gamma(a_1+a_2+a_3+a_4)}{\Gamma(a_1)\Gamma(a_2)\Gamma(a_3)\Gamma(a_4)}\theta_1^{a_1-1}\theta_2^{a_2-1}\theta_3^{a_3-1}(1-\theta_1-\theta_2-\theta_3)^{a_4-1} \quad (4.2)$$

where $\boldsymbol{\theta} = (\theta_1, \theta_2, \theta_3)$

Now the posterior distribution of the parameters θ_1, θ_2 and θ_3 using (4.2) and the likelihood (4.1) is:

$$p(\theta_1,\theta_2,\theta_3|\mathbf{x}) = \frac{\theta_1^{n_1+a_1-1}\theta_2^{n_2+a_2-1}\theta_3^{n_3+a_3-1}(1-\theta_1-\theta_2-\theta_3)^{n_4+a_4-1}}{K(\theta_1+\theta_2)^{r_{12}}(\theta_2+\theta_3)^{r_{23}}(\theta_1+\theta_3)^{r_{13}}(1-\theta_2-\theta_3)^{r_{14}}(1-\theta_1-\theta_3)^{r_{24}}(1-\theta_1-\theta_2)^{r_{34}}} \quad (4.3)$$

where K is normalizing constant and $\theta_1, \theta_2, \theta_3 \geq 0$, $\theta_1+\theta_2+\theta_3 < 1$.

The marginal posterior densities for the parameters are obtained by using (4.3). The graphs of the marginal posterior densities of θ_1, θ_2, θ_3 and θ_4 are presented in Fig4.1.

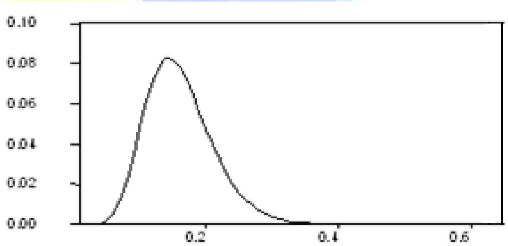
We simulate data (given in Table 4.1) for the paired comparison of four treatments assuming the Bradley-Terry model with the parameters values as: θ_1=0.16, θ_2=0.23, θ_3=0.31, θ_4=0.30 and the number of comparisons $(r_{ij} = 10)$:

Table 4.1:-Simulated Data for m=4

Pairs	(1,2)	(1,3)	(1,4)	(2,3)	(2,4)	(3,4)
n_{ij}	5	4	2	6	4	7
n_{ji}	5	6	8	4	6	3
r_{ij}	10	10	10	10	10	10

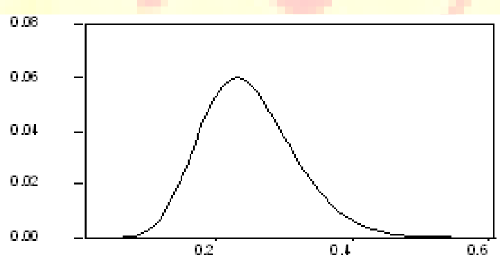
Posterior (Marginal) Densities for the Parameters of the Bradley-Terry Model

$p(\theta_1|\mathbf{x})$

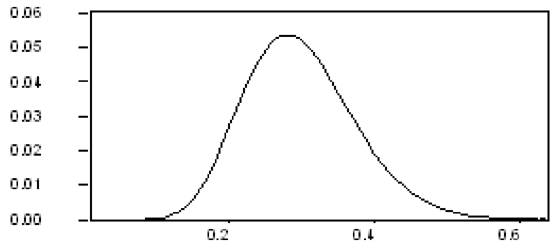


θ_1

$p(\theta_2|\mathbf{x})$



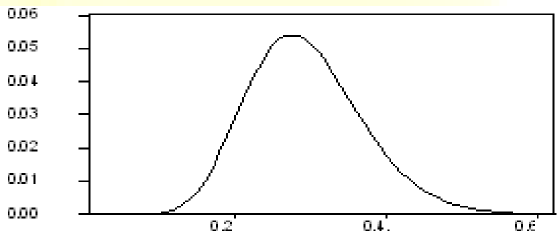
θ_2



$p(\theta_3|\mathbf{x})$

θ_3

$p(\theta_4|\mathbf{x})$



θ_4

Figure 4.1

4.1.1 Posterior Estimates

Using the posterior distribution (4.3), two programs are designed in SAS package to find the posterior estimates: means and modes. Having run the programs, the obtained results are presented in Table 4.2 for the data set given in Table 4.1

Table 4.2:-Posterior Estimates

Parameters	θ_1	θ_2	θ_3	θ_4
Posterior Means	0.1602	0.2494	0.2974	0.2930
Posterior Modes	0.1505	0.2350	0.2851	0.3294

It is observed that the joint posterior mode is similar to the posterior means and indicate the similar ranking of the treatments i.e. $T_3 \to T_4 \to T_2 \to T_1$.

4.1.2 Preference Probabilities

As we know that preference probabilities ψ_{ij} are the probabilities of preferring treatment T_i to treatment T_j. For determining ψ_{ij} and $\psi_{ji} = 1 - \psi_{ij}$, we use the values of the posterior means: θ_1 =0.1602, θ_2 =0.2494, θ_3 =0.2974 and θ_4 =0.2930. The preference probabilities are shown in Table 4.3:

Table 4.3:- Preference Probabilities

Preference Prob	ψ_{12}	ψ_{13}	ψ_{14}	ψ_{23}	ψ_{24}	ψ_{34}
Values	0.4912	0.3501	0.3036	0.4561	0.4599	0.5538
Preference Prob	ψ_{21}	ψ_{31}	ψ_{41}	ψ_{32}	ψ_{42}	ψ_{43}
Values	0.5088	0.6499	0.6964	0.5439	0.5401	0.4462

4.1.3 Posterior Probabilities of Hypotheses

The following hypotheses H_{ij} and $\bar{H}_{ij}$ (i<j=1,2,3,4) are compared:

$H_{ij} : \theta_i > \theta_j$ and $\bar{H}_{ij} : \theta_j \geq \theta_i$.

The posterior probability p_{ij} for H_{ij} is $p_{ij} = p(\theta_i > \theta_j)$ and $q_{ij} = 1 - p_{ij}$ is the posterior probability for $\bar{H}_{ij}$. The posterior probability { p_{12} } for H_{12} is obtained as:

$$p_{12} = p(\theta_1 > \theta_2) = p(\theta_1 - \theta_2 > 0) = p(\phi > 0)$$

$$= \int_{\phi=0}^{1} \int_{\xi=\phi}^{\frac{1+\phi}{2}} \int_{\theta_3=0}^{1-2\xi+\phi} \frac{\xi^{n_1+a_2-1}(\xi-\phi)^{n_2+a_2-1}\theta_3^{n_3+a_3-1}(1-2\xi+\phi-\theta_3)^{n_4+a_4-1}}{KD} d\theta_3 d\xi d\phi \quad (4.4)$$

where $\phi = \theta_1 - \theta_2$ and $\theta_1 = \xi$, K is normalizing constant and D is defined as:

$$D = (2\xi-\phi))^{r_{12}}((\xi-\phi)+\theta_3)^{r_{23}}(\xi+\theta_3)^{r_{13}}(1-(\xi-\phi)-\theta_3)^{r_{14}}(1-\xi-\theta_3)^{r_{24}}(1-2\xi-\phi)^{r_{34}}.$$

By SAS package, we get posterior probabilities shown in Table 4.4:

Table 4.4:- Posterior Probabilities

Pairs(i,j)	(1,2)	(1,3)	(2,3)	(1,4)	(2,4)	(3,4)
p_{ij}	0.1447	0.0073	0.3274	0.0077	0.3398	0.4981
q_{ij}	0.8553	0.9927	0.6726	0.9923	0.6602	0.5019

The hypothesis with greater probability will be accepted. Let s= $\min(p_{ij}, q_{ij})$, if s>0.1 the decision is inconclusive. Here $\bar{H}_{13}$ and $\bar{H}_{14}$ are accepted with high probability and all other hypotheses seem inconclusive.

4.1.4 Predictive Probabilities

Let the predictive probability $P_{(12)}$ that treatment T_1 would be preferred to treatment T_2 in a future single comparison of two treatments is{using (2.1) and (4.9)} is:

$$P_{(12)} = P(T_1 > T_2) = \int_{\theta_1=0}^{1} \int_{\theta_2=0}^{1-\theta_1} \int_{\theta_3=0}^{1-\theta_1-\theta_2} \psi_{12}\, p(\theta_1, \theta_2, \theta_3 | \mathbf{x})\, d\theta_3 d\theta_2 d\theta_1 \qquad (4.5)$$

Using SAS package, the predictive probabilities are given as follows in Table 4.5:

Table 4.5:- Predictive Probabilities

Pairs(i,j)	(1,2)	(1,3)	(2,3)	(1,4)	(2,4)	(3,4)
$P_{(ij)}$	0.3933	0.3530	0.4570	0.3564	0.4607	0.3933
$P_{(ji)}$	0.6067	0.6470	0.5430	0.6436	0.5393	0.6067

We would always expect that the predictive probabilities are near to 0.5.

5.1 <u>Appropriateness of the Model</u>

For the data given in Table 4.1, we consider chi-square goodness of fit test for the model fitting. The expected numbers of preferences are calculated by multiplying preference probabilities (given in Table 4.2) with number of comparisons (r_{ij}=10). The following chi-squared test is used to test the appropriateness of the model:

$$\chi^2 = \sum_{i<j=1}^{m} \frac{\{(n_{ij} - \hat{n}_{ij})^2}{n_{ij}} \quad \text{with} \quad \frac{(m-1)(m-2)}{2} \text{ degrees of freedom.}$$

Table 5.1:- Observed and Expected Number of Preference

Pairs	(1,2)		(1,3)		(1,4)		(2,3)		(2,4)		(3,4)	
	n_{12}	$n_{2.}$	n_{13}	n_{31}	n_{14}	n_{41}	n_{23}	n_{32}	n_{24}	n_{42}	n_{34}	n_{43}
Observed	5	5	4	6	2	8	4	6	4	6	7	3
Expected	5	5	4	6	3	7	5	5	5	5	6	4

We find χ_1^2=1.6929 with p-value 0.6385 indicates that there is no evidence that the model does not fit the data.

References

- Adam, E. S. (2005). Bayesian Analysis of Linear Dominance Hierarchies. *Animal Behavior*. Vol 69. issue, 5. 1191-1201.
- Aslam, M. (2003). An application of prior predictive distribution to elicit the Prior density. *Journal of Statistical Theory and Applications*, Vol.2, 1, 70-83.
- Bradley, R. A., and Terry. M.E. (1952). Rank Analysis of Incomplete Block Design:I. The Method of Paired Comparisons. *Biometrika* 39, 324-345.
- Causeur D. & Husson F. (2005). A 2 dimensional extension of the Bradley-Terry model for paired comparisons. *Journal of Statistical Planning and Inference*. 135. pp. 245-259**.**
- Chaloner, K. M. and Duncan, G. T. (1983). Assessment of a beta prior distribution: PM elicitation. *The Statistician* 32, ½,174-180.
- David, H. A., (1988). The method of paired comparisons, 2nd edn., Charles Griffin and Company-London.

- Davidson, R. R. and Solomon, D. L. (1973). A Bayesian Approach to Paired Comparison Experimentation. *Biometrika*, 60, 447-487.
- Jian, B. (2007). Similarity testing using paired comparison method, *Food Quality and Preference*, Volume 18, Issue 3, 500-507.
- Kadane, J. B. (1966). Some equivalence classes in paired comparisons. *The Annals of Mathematical statistics*, vol.37, No.2, 488-494.
- Kadane, J. B., Dickey, J. M., Winkler, R. L., Smith, W. and Peter, S.C. (1980). Interactive elicitation of opinion for a normal linear model. *JASA.* **75**, 845-854.
- Leonard, T. (1977). An alternative Bayesian approach to the Bradley-Terry model for paired comparisons. *Biometrics* 33, 121-132.
- Kim, D-H and Kim, H-J. (2004). A Bayesian Approach to Paired Comparison of Several Products of Poisson Rates. Proceedings of the Autumn Conference, Korean Statistical Society.
- Miranda. E, Bourque .P, Abran. A(2009) *Information and Software Technology* , Volume 51, Issue 9, September 2009, 1327-1337.
- Winkler, R. L. (1980) Prior Information, Predictive Distributions and Bayesian Model
- Building. Bayesian analysis in econometrics and statistics, Zellner, A (Ed.) North- Holland, Amsterdam. 95-109.
- Zermelo, E. (1929) Die Berechnung der Turnier-Ergebnisse als ein Maximumproblem der Warscheinlichkeitsrechnung. *Math. Zeit*., 29, 436-460.