# Bayesian Analysis for the Paired Comparison Model with Order Effects (Using Non-informative Priors)

Ghausia Masood Gilani College of Statistical and Actuarial Sciences University of the Punjab, Q.A. Campus, Lahore, Pakistan gilani\_pu@yahoo.com

Nasir Abbas Department of Statisics, Government Postgraduate College Jhang, Pakistan nabbasgcj@yahoo.com

#### Abstract

Sometimes it may be difficult for a panelist to rank or compare more than two objects or treatments at the same time. For this reason, paired comparison method is used. In this study, the Davidson and Beaver (1977) model for paired comparisons with order effects is analyzed through the Bayesian Approach. For this purpose, the posterior means and the posterior modes are compared using the noninformative priors.

**Keywords:** Paired comparison method; The Davidson and Beaver model; Prior distribution; Noninformative prior; Jeffreys prior; Uniform distribution; Posterior distribution.

#### 1. Introduction

Sometimes- it may be difficult for a panelist to simultaneously rank or compare more than two treatments (objects, items, options, stimuli etc.) specially when differences between treatments are small or the criteria are rather subtle. For this reason, the paired comparison data is regarded as more reliable and can be obtained more readily from panelists. In a paired comparison trial, a panelist is given a pair of treatments and is asked to pick the better one with respect to a given attribute. This process is repeated for all pairs of the treatments under study.

This method is widely used in industry for assessing customer preference and designing products using trained panelists. For example, in taste testing it is often difficult for a panelist to cope with more than two tastes at the same time and the introduction of a third may be confusing. A good review of the paired comparison models including their analysis is given by Bradley (1976). David (1988) has a detailed survey of the literature and references concerning the method of paired comparisons. Further literature can be seen in Augustin (2004), David, (2004), Hatzinger *et. al.* (2004) and the references cited therein. Some recent developments include Abbas and Aslam (2009) who develop a paired comparison model based on Cauchy distribution and analyze it via the Bayesian approach. Abbas and Aslam (2010) perform Bayesian analysis of the chi-square models suggested by Stern (1990).

#### 2. The Davidson and Beaver Model

Consider a paired comparison trial with a set of *m* treatments  $T_1, T_2, ..., T_m$ . Each pair formed from the set of *m* treatments is ranked  $r_{ij}$  times by being presented to a respondent who is asked to indicate a preference for one treatment of the pair. It is assumed that the responses to the treatments can be described in terms of an underlined continuum on which the relative *worth* of the treatments can be located. We denote the *worth* as an index of relative merit of the treatment  $T_i$  by  $\theta_i$ ,  $\theta_i > 0$ , where, without loss of generality,  $\sum \theta_i = 1$ . A model can be based on the idea that when a panelist is confronted with treatment  $T_i$ , he responds with an unconscious or latent variable  $X_i$ . The assumed mechanism is that he prefers treatment  $T_i$  to treatment  $T_j$  if  $X_i > X_j$ .

Bradley and Terry (1952) propose a model of paired comparisons for the trial mentioned above. They assume that the difference between two latent variables i.e.,  $(X_i - X_j)$ , has a logistic (squared hyperbolic secant) density with location parameter  $(\ln \theta_i - \ln \theta_j)$ . Now 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, is defined as:

$$\psi_{ij} = P(X_i > X_j \mid \theta_i, \theta_j) = 1/4 \int_{-(\ln \theta_i - \ln \theta_j)}^{\infty} \sec h^2(y/2) dy = \frac{\theta_i}{\theta_i + \theta_j}, \ i(\neq j) = 1, 2, ..., m$$
(1)

The Bradley-Terry model is defined by (1).

Davidson and Beaver (1977) propose a modification of the Bradley -Terry model to account for the effect of the order of presentation of the treatments within a pair. A multiplicative order effect is suggested as an alternative to the additive order effect proposed by Beaver and Gokhale (1975). An important feature of the Bradley-Terry model is that the values  $\ln \theta_1$ ,  $\ln \theta_2$ , ...,  $\ln \theta_m$  can be used to represent the merits of the treatments under study on a linear scale. Thus it is natural to assume that the logarithms of the worths, as supposed to the worths themselves, are affected additively by the order of presentation. This is equivalent to assume that when treatment  $T_i$  and  $T_j$  appear together in a pair, their relative worths are subject to a multiplicative within-pair order effect  $\gamma_{ij}$ . It is assumed that  $\gamma_{ij} = \gamma_{ji}$ , i.e., the within-pair order effect depends only on the treatments pair. The resulting preference probabilities for the ordered pair ( $T_i$ ,  $T_j$ ) are given by

$$\psi_{ij} = 1/4 \int_{-(\ln\theta_i - \ln\theta_j) + \gamma_{ij}}^{\infty} \sec h^2(y/2) dy = \frac{\theta_i}{\theta_i + \gamma_{ij}\theta_j}, \qquad i(\neq j) = 1, 2, ..., m,$$
(2)

and  $\psi_{ji} = \frac{\gamma_{ij}\theta_j}{\theta_i + \gamma_{ij}\theta_j}$ . Here  $\gamma_{ij} > 0$ . When  $\gamma_{ij} = 1$ , there is no order effect and the model

yields the Bradley-Terry model (1). When  $\gamma_{ij} > 1$ , the worth of the treatment presented second becomes inflated while if  $\gamma_{ij} < 1$ , the treatment presented first gains an advantage. The case  $\gamma_{ij} = \gamma$ , for all (*i*, *j*) is of interest because of the

considerable reduction in the number of parameters required to specify the model. The multiplicative model (2) arises naturally in the setting of the linear model, (David 1988). Suppose that an individual experiences sensations  $X_i$  and  $X_j$  when presented the pair (i, j), and that the response  $i \rightarrow j$  (*i* preferred to *j*) results when  $X_i > X_j$  is interpreted to mean that the sensation to treatment  $T_i$  comes closer to the ideal sensation than that for treatment  $T_j$ . It is noted by Bradley (1953) that when the difference  $X_i - X_j$  has a logistic distribution with location parameter  $(\ln \theta_i - \ln \theta_j)$ , one obtains the preference probabilities of the Bradley-Terry model. Now suppose that for the ordered pair (i, j), the response  $i \rightarrow j$  results when  $X_i - X_j > \tau_{ij}$  where  $\tau_{ij}$  is interpreted as a shift on the sensation scale which arises because of the order of presentation. Using the logistic distribution for  $X_i - X_j$  one obtains the preference probabilities given by the multiplicative model (2) with  $\tau_{ij} = \ln \gamma_{ij}$ .

So far it has been assumed that each response to a pair of treatments consists of an indication of preference for one member of the pair.

#### 3. Notations and Likelihood Function for the Model

Let  $w_{ijk}(1)$  and  $w_{ijk}(2)$  be the random variables associated with the rank of the treatment *i* in the *k*th repetition of the treatment pair  $(T_i, T_j)$ ,  $i (\neq j) = 1, 2, 3, ..., m$ ,  $k = 1, ..., r_{ij}$ . They are defined as:

 $w_{ijk}(1) = 1$  or 0 according as the treatment  $T_i$  is preferred to treatment  $T_j$  or not in the *k*th repetition of comparison.

 $w_{ijk}(2) = 1$  or 0 according as the treatment  $T_j$  is preferred to treatment  $T_i$  or not in the *k*th repetition of comparison.

 $w_{ij}(1) = \sum_{k} w_{ijk}(1)$  = The frequency of preference for the treatment

presented first (treatment *i*).

 $w_{ij}(2) = \sum_{k} w_{ijk}(2)$  = The frequency of preference for the treatment

presented second (treatment *j*)

 $r_{ij}$  = The number of timed treatment  $T_i$  is compared with treatment  $T_j$ and  $r_{ij} = w_{ij}(1) + w_{ij}(2)$ 

$$N = \sum_{i} \sum_{j \neq i} r_{ij}$$
 = The total number of comparisons  
$$w_i = \sum_{j} \left\{ w_{ij}(1) + w_{ij}(2) \right\}$$
 = The total number of wins of the treatment  $T_i$ 

Now we derive the likelihood function of the data for the Davidson and Beaver model stated in (2). We put constraint on the treatment parameters of the model that they are positive and they sum to unity. These conditions ensure that the parameters are well defined and identifiable.

The probability of the observed result in the kth repetition of the pair ( $T_i$ ,  $T_j$ ) is

$$P_{ijk} = \left(\frac{\theta_i}{\theta_i + \gamma \theta_j}\right)^{w_{ijk}(1)} \left(\frac{\gamma \theta_j}{\theta_i + \gamma \theta_j}\right)^{w_{ijk}(2)}$$
(3)

Hence, the likelihood function of the observed outcome x (where x represents the data  $(w_{ijk}(1), w_{ijk}(2))$  of the trial is

$$L(x;1,2) = \prod_{k=1}^{r_{ij}} \prod_{i(\neq j)=1}^{m} P_{ijk} = \frac{\gamma^{\sum_{i=1}^{m} \sum_{j\neq i}^{m} w_{ij}(2)} \prod_{i=1}^{m} \theta_{i}^{\sum_{i=1}^{m} w_{ij}(1)} \prod_{j=1}^{m} \theta_{j}^{\sum_{j=1}^{m} w_{ij}(1)} \prod_{i(\neq j)=1}^{m} k_{ij}}{\prod_{i=1}^{m} \prod_{j=1}^{m} (\theta_{i} + \gamma \theta_{j})^{w_{ijk}(1) + w_{ijk}(2)}},$$
(4)

where  $0 < \theta_i < 1, i = 1, 2, ..., m$ ,  $k_{ij} = r_{ij} / \{ w_{ij}(1)! w_{ij}(2)! \}$ ,  $\sum_{i=1}^m \theta_i = 1 \text{ and } \gamma > 0$ . Here  $\gamma$  is

the order effect parameter and  $\theta_1, \theta_2, ..., \theta_m$  are the treatment parameters.

#### 4. The Choice of Prior Distribution

Bayesian analysis is a statistical procedure, which endeavors to estimate parameters of an under lying distribution based on the observed distribution. We begin with the derivation of the prior distribution of parameters, include an assessment of the likelihoods function of the sample observations derived from the distribution identified by the parameters and then merge them to yield a posterior distribution of the parameters of interest. Then we base the entire parametric inference on the very posterior distribution derived for the parameters of interest conditional upon the data. In practice, it is common to assume a uniform distribution over the appropriate range of values of the parameters for the prior distribution. Adams (2005) throws light on the advantages of the Bayesian approach in an explained way.

Prior distribution quantifies information about parameter prior to any data being gathered. The prior which expresses specific definite information about a random variable is the informative prior. But in some cases, such as multiparameter situations, it becomes difficult to formalize any available prior information in to a distribution. In such cases when little prior information is known or prior elicitation is difficult, analysis is done by choosing a prior which reflect little prior information. These priors are known as non-informative priors. A prior distribution is non-informative if it is *flat* relative to the likelihood function. Thus a prior distribution of parameter of interest and is dominated by the likelihood, that is, it does not change very much over the region in which the likelihood is appreciable and does not assume large values outside the range. A prior which has these properties is said to be a locally uniform prior. Other names for non-informative priors are reference priors, vague priors, ignorant priors or flat priors.

Many approaches for the choice of a non-informative prior have been given. One way of eliminating the subjectivity in the choice of prior is to use a flat or diffuse prior distribution that is uniform across all possible values of the parameter. Such a non-informative diffuse prior is simply a constant, i.e.,  $p(\theta) = c$ , for  $\theta$  belonging to the parametric space. With a diffuse prior, the posterior is just a constant *c* times the likelihood, i.e.

$$p(\theta \mid x) = p(\theta)L(\theta \mid x) \implies p(\theta \mid x) = cL(\theta \mid x) \implies p(\theta \mid x) \propto L(\theta \mid x)$$

Another approach is using the Jeffreys prior. It satisfies the local uniformity property for non-informative priors. It is the prior based on the Fisher information matrix.

#### 5. Reference (Jeffreys) Prior for the Parameters of the Model

A non-informative prior has been suggested by Jeffreys (1946, 1961) which is frequently used in the situation where one does not have much information about the parameters. It is defined as the density of the parameters proportional to the square root of the determinant of the Fisher's Information Matrix.

Let  $p(\mathbf{x}|\theta)$  denotes the density of **x** given  $\theta$ . The Fisher information is

$$I(\theta) = -E\left[\frac{\partial^2 \log p(\mathbf{x} \mid \theta)}{\partial \theta^2}\right]$$

If  $\theta$  is a p×1 vector then

$$I(\ ) = -E\left[\frac{\partial^2 \log p(\mathbf{x} \mid \ )}{\partial \theta_i \partial \theta_j}\right]$$

and thus I() is a  $p \times p$  matrix. The Jeffreys prior is defined as the determinant of the Fisher information matrix, i.e.,  $p() \propto \sqrt{\det\{I()\}}$ .

## Properties of the Jeffreys' Prior

The Jeffreys Prior shows many nice properties that make it an attractive non-informative prior. The Jeffreys prior has invariance property with regards to its one to one (1-1) transformation of the parameter in the sense that we get consistent answers in any parameterization.

Bernardo (1979) shows that the Jeffreys prior is the appropriate reference prior if there are no parameters that are regarded as the nuisance parameters and the joint posterior distribution of all the parameters is asymptotically normal. Another important aspect of the prior is that it is not effective by a restriction on the parametric space.

If the likelihood function (4) belongs to the exponential family and it follows from the regularity conditions of Johnson and Ladalla (1979) that the posterior distribution is asymptotically normal. Here no parameter is regarded as 'nuisance' so the Jeffreys prior is the appropriate choice of non-informative prior and hereafter will be called the reference (Jeffreys) Prior.

Let us consider the case for the number of treatments *m*=2. The likelihood function for the parameters  $\gamma$ ,  $\theta_1$  and  $\theta_2$  of the Davidson and Beaver model is proportional to

$$l(x;\phi) = \frac{\gamma^{w_{12}(2)+w_{21}(2)}\theta^{w_{12}(1)+w_{21}(2)}(1-\theta)^{w_{21}(1)+w_{12}(2)}}{(\theta+\gamma(1-\theta))^{w_{12}(1)+w_{21}(2)}}$$
$$= \frac{\gamma^{w_{12}(2)+w_{21}(2)}\theta^{w_{12}(1)+w_{21}(2)}(1-\theta)^{w_{21}(1)+w_{12}(2)}}{\{\theta+\gamma(1-\theta)\}^{r_{12}}\{(1-\theta)+\gamma\theta\}^{r_{21}}}$$

Where  $\phi = (\gamma, \theta)$ ,  $\theta_1 = \theta$ ,  $\theta_2 = 1 - \theta$  and  $r_{ij} = w_{ij}(1) + w_{ij}(2)$ .

Using the MAPLE package, the reference (Jeffreys) prior is:

$$p(\phi) \propto \frac{1}{\{\theta - \gamma(1 - \theta)\}\{(1 - \theta) - \gamma\theta\}}$$
  
$$\Rightarrow p(\phi) = \frac{1}{K\{\theta - \gamma(1 - \theta)\}\{(1 - \theta) - \gamma\theta\}}, \quad 0 \le \theta \le 1, \ 0 \le \gamma \le \infty$$

where  $K = 5.349 \times 10^{-29}$ .

## 6. Bayesian Analysis of the Model

Using the data given in Table 1, the Bayesian analysis of the model with order effect for two treatments with equal number of comparisons for each pair  $r_{ij} = 50$ , (i, j = 1, 2), is presented using the non- informative priors: the uniform and the reference (Jeffreys) priors.

## Table 1: Data with Ordered Effects

Pairs	r <sub>ij</sub>	<i>w<sub>ij</sub></i> (1)	<i>w<sub>ij</sub></i> (2)
(1, 2)	50	14	36
(2, 1)	50	32	18

## 6.1 The Posterior Distributions

The (joint) posterior distribution using the *uniform prior*  $\{p(\theta, \gamma) \propto 1\}$ , for the treatment parameter  $\theta_1$  (with constraint  $\theta_2 = 1 - \theta_1$ ) and the order effect parameter  $\gamma$  is

$$p(\theta_{1}, \gamma \mid \mathbf{x}) = \frac{\gamma^{w_{12}(2) + w_{21}(2)} \theta_{1}^{w_{12}(1) + w_{21}(2)} (1 - \theta_{1})^{w_{21}(1) + w_{12}(2)}}{K \left\{ \theta_{1} + \gamma (1 - \theta_{1}) \right\}^{r_{12}} \left\{ (1 - \theta_{1}) + \gamma \theta_{1} \right\}^{r_{21}}}, \ \mathbf{0} \le \theta \le 1, \ \mathbf{0} \le \gamma \le 1$$

where  $K = 6.7355 \times 10^{-29}$  is the normalizing constant. Here we may replace  $\theta_1$  by  $\theta$  for simplicity.

The (marginal) posterior densities of the parameters  $\theta_1$  and  $\gamma$  are:

$$p(\theta_1 \mid \mathbf{x}) = \int_0^\infty p(\theta_1, \gamma \mid \mathbf{x}) d\gamma, \ \mathbf{0} \le \theta_1 \le \mathbf{1}$$
(5)

$$p(\gamma \mid \mathbf{x}) = \int_{0}^{1} p(\theta_{1}, \gamma \mid \mathbf{x}) d\theta_{1}, \ \mathbf{0} \le \gamma \le$$
(6)

Similarly the (marginal) posterior densities of the parameters  $\gamma$  can be derived. The (joint) posterior distribution using the Jeffreys prior for the treatment parameters  $\theta_1$  (with constraint  $\theta_2 = 1 - \theta_1$ ) and the order effect parameter  $\gamma$  is

where K = 5.3549 × 10<sup>-29</sup> is the normalizing constant. The (marginal) posterior densities of the parameters  $\theta_1$  and  $\gamma$  may be found using (5) and (6).

#### 6.2 The Posterior Estimates

The posterior means and joint posterior modes of the parameters are considered to be the estimates of the parameters. The posterior means of the parameters using uniform prior and Jeffreys priors are evaluated using the quadrature method for the data set given in Table 1. We evaluate the expressions

$$E(\theta_1 | \mathbf{x}) = \int_{0}^{1} \theta_1 p(\theta_1 | \mathbf{x}) d\theta_1$$
 and  $E(\gamma | \mathbf{x}) = \int_{0}^{\infty} \gamma p(\gamma | \mathbf{x}) d\gamma$  respectively to find the mean

estimates the worth and order effect parameters. Using the uniform prior, the values of the posterior means of the parameters  $\theta_1$ ,  $\theta_2$  and  $\gamma$  are evaluated to be 0.31972, 0.68028 and 1.296670, whereas using the Jeffreys prior, these estimates are 0.32050, 0.67950 and 1.230970. Here it can be seen that the posterior means of the parameters evaluated using the uniform prior are very close to those produced using the Jeffreys prior.

The estimates of the parameters of interest which maximize the posterior density are termed as the joint posterior modes. These are found by solving the equations obtained by equating to zero the first partial derivatives of the logarithm of the likelihood function with regards to the unknown parameters. That is we solve:

$$\frac{\partial \ln L(\phi)}{\partial \theta_1} = 0, \quad \frac{\partial \ln L(\phi)}{\partial \theta_2} = 0 \text{ and } \frac{\partial \ln L(\phi)}{\partial \gamma} = 0$$

The modal estimates using the uniform prior are found for the parameters  $\theta_1$ ,  $\theta_2$ and  $\gamma$  by executing a computer program designed in the SAS package. The values of the posterior joint mode of the vector ( $\theta_1$ ,  $\theta_2$ ,  $\gamma$ ) are obtained to be ( $\theta_1 = 0.318665$ ,  $\theta_2 = 0.681335$ ,  $\gamma = 1.22676$ ). Similarly the modal estimates evaluated using the Jeffreys prior are found by running another SAS program and the values of the joint posterior mode of the parameter vector ( $\theta_1$ ,  $\theta_2$ ,  $\gamma$ ) are obtained to be ( $\theta_1 = 0.319267$ ,  $\theta_2 = 0.680733$ ,  $\gamma = 1.145057$ ). Here we observe that the resulting estimates produced using both the non-informative priors – the Jeffreys and the uniform priors – are also very close.

## 7. Conclusions

Here we perform the Bayesian analysis of the Davidson Beaver model for paired comparisons based on the non-informative priors – the Jeffreys as well as the uniform priors. We derive the means and the modal estimates of the model parameters for an illustration based on the observed data. From the results it reveals that the posterior estimates obtained using the two types of the non-informative priors are very similar, which gives confidence in using either of the priors. The results also exhibit the robustness with respect to the choice of non-informative priors: the Jeffreys and the uniform. This means the simpler uniform prior is a reasonable option in this case.

## References

- 1. Abbas, N. and Aslam, M. (2009). Prioritizing the Items through Paired Comparison Models, A Bayesian Approach. *Pakistan Journal of Statistics*, 25(1), 59-69.
- 2. Abbas, N. and Aslam, M. (2010). Bayesian Analysis of Gamma Models for Paired Comparisons. *Pakistan Journal of Statistics*, (in press).
- 3. Adams, E.S. (2005). Bayesian analysis of linear dominance hierarchies. *Animal Behavior*, 69, 1191-1201.
- 4. Augustin, T. (2004). Bradley-Terry-Luce models to incorporate within-pair order effects; Representation and uniqueness theorems. *British Journal of Mathematical and Statistical Psychology*, 57, 281-294.
- 5. Beaver, R. J. and Gokhale, D.V (1975). A model to incorporate within-pair order effects in paired comparisons. *Comm. Statist,* 4, 923-939.
- 6. Bernardo, J. M. (1979). Reference Posterior Distributions for Bayesian Inference (with discussion), *Journal of the Royal Statistical Society Series B*, 41, 113-147.
- 7. Bradley, R. A and Terry, M. E. (1952). Rank analysis of incomplete Block Designs I. The method of paired comparisons. *Biometrika*, 39, 324-45.
- 8. Bradley, R. A. (1976). Science, Statistics and paired comparisons. *Biometrics*, 32, 213-232.
- 9. Bradley, R.A. (1953). Some Statistical methods in taste testing and Quality Evaluation. *Biometrics*, 9, 22-38.
- 10. David, H. A. (1988). *The method of paired comparisons*. 2<sup>nd</sup> ed. London: Griffin.
- 11. David, R. H. (2004. MM Algorithms for generalized Bradley-Terry models. *The Annals of Statistics*, 32, 384-406.

- 12. Davidson, R. R. and Beaver, R. J. (1977). On extending the Bradley-Terry model to incorporate within-pair order effects. *Biometrics*, 33, 693-702.
- 13. Hatzinger, Reinhold: Francis, Brian, J. (2004). Fitting paired comparison models in R. Research Report Series / Department of Statistics and Mathematics Nr. 3. May 2004.
- 14. Jeffreys, H. (1946). An Invariant Form for the Prior Probability Estimation Problems. *Proceeding of the Royal Society of London, Series A*, 186, 453-461.
- 15. Jeffreys, H. (1961). *Theory of Probability*, Oxford UK: Claredon Press.
- 16. Johnson, R. A. and Ladalla, J. N. (1979). The large sample behavior of posterior distributions when sampling from multiparameter exponential family models and allied results, *Sankhya*, *Ser. B*, 41, 196–215.
- 17. Stern, H. (1990a). A Continuum of Paired Comparison Models. *Biometrika*, 77(2), 265-273.