

# Lindley-Shannon Information for Comparison of Priors Under Paired Comparisons Model

Muhammad Aslam  
Department of Basic Sciences  
Riphah International University, Islamabad, Pakistan  
aslamsdqu@yahoo.com, m.aslam@riphah.edu.pk

Syed Haider Shah  
Department of Statistics  
University of Balochistan, Quetta, Pakistan  
mrhaidershah@yahoo.com

## Abstract

One of the main differences between classical statistics and Bayesian statistics is that the latter can utilize prior information in a formal way. This information can be quantified in terms of a probability distribution which is known as the prior distribution. If there is no relevant prior information available then there are ways to derive a 'non-informative' prior distribution. In this study, the informative and non-informative prior distributions for the parameters of the Bradley-Terry model are compared through the Lindley-Shannon information.

**Keywords:** Prior distribution, non-informative prior, informative prior, paired comparison method, the Bradley-Terry model.

## 1. Introduction

For a Bayesian analysis, prior information in terms of probability distribution is used. If there is no relevant prior information available then there are ways to derive a 'non-informative' prior distribution. In Section 2, the Bradley-Terry model for paired comparisons is defined. Section 3 presents the informative prior distribution for the parameters. Non-informative prior distribution from Jeffreys rule is derived in Section 4. Comparison of non-informative and informative priors through the Lindley-Shannon information {Lindley (1956), Shannon (1948)} is discussed in Section 5. Conclusion is drawn in the last Section 6. The graphs of non-informative priors are also drawn for this purpose in this study.

## 2. The Bradley-Terry Model for Paired Comparison

Sometimes it may be difficult for a panelist to rank or compare more than two objects or treatments ( $m > 2$ ) at the same time especially when differences between objects are small or the criteria are rather subtle ( $m = 2$ ). For this reason paired comparison data is sometimes regarded as more reliable and can be obtained more readily from panelists. In a paired comparison trial a panelist is given a pair of objects and asked to say 'with respect to a given attribute' which is preferable. This is repeated for several pairs. This method is widely used in industry for assessing customer preference and designing products using trained panelists. David (1988) has a detailed survey of the literature and references concerning the method of paired comparisons. Further literature can be seen in Gilani and Abbas (2008), Abbas (2010) and Altaf, et. al (2011).

Let consider a paired comparison trial with 'm' treatments or objects  $T_1, T_2, \dots, T_m$ . There are ' $m(m-1)/2$ ' all possible pairs of 'm' treatments. Each pair  $(T_i, T_j)$ , ( $i < j, 1 \leq i, j \leq m$ ) is ranked  $r_{ij}$  times independently. Also let  $\theta_1, \theta_2, \dots, \theta_m$  be the 'true' ratings (or preferences) of 'm' treatments on a subjective continuum. A model can be based on the idea that when a panelist is confronted with treatment  $T_i$ , she responds with an 'unconscious' or 'latent' variable  $X_i$ . The assumed mechanism is that she 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. The Bradley-Terry model implies that difference between two latent variables  $(X_i - X_j)$  has a logistic (squared hyperbolic secant) density with location parameter  $(\ln \theta_i - \ln \theta_j)$ . If  $\psi_{ij}$  denotes the probability  $\mathbf{P}(X_i > X_j | \theta_i, \theta_j)$  that treatment  $T_i$  is preferred to treatment  $T_j$  ( $i \neq j$ ) when the treatments  $T_i$  and  $T_j$  are compared then

$$\begin{aligned} \psi_{ij} &= \frac{1}{4} \int_{-(\ln \theta_i - \ln \theta_j)}^{\infty} \operatorname{sech}^2(y/2) dy && i \neq j, i, j = 1, 2, \dots, m. \\ &= \frac{\theta_i}{\theta_i + \theta_j}. \end{aligned} \tag{1}$$

The Bradley-Terry model is defined by (1)

**The following notations are used in the comparisons:**

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

It is to be noted that  $n_{ijk} + n_{jik} = 1$  and  $r_{ij} = n_{ij} + n_{ji} = r_{ji}$ .

We put constraints on the parameters of the model that they are positive and they sum to unity, these conditions ensure that the parameters are well defined i.e. identifiable.

The probability of the observed result in the  $k$ 'th repetition of the pair  $(T_i, T_j)$  is:

$$P_{ijk} = \left( \frac{\theta_i}{\theta_i + \theta_j} \right)^{n_{i,\bar{k}}} \left( \frac{\theta_j}{\theta_i + \theta_j} \right)^{n_{j,\bar{k}}} \tag{2}$$

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

$$l(\mathbf{x}; \theta_1, \dots, \theta_m) = \prod_{1 \leq i < j \leq m} \prod_{k=1}^{r_{ij}} P_{ijk} = \prod_{i(<j)=1}^m \binom{r_{ij}}{n_{ij}} \frac{\theta_i^{n_{ij}} \theta_j^{n_{ji}}}{(\theta_i + \theta_j)^{r_{ij}}}$$

$$= \prod_{i=1}^m \theta_i^{n_i} \prod_{i < j=1}^m \binom{r_{ij}}{n_{ij}} (\theta_i + \theta_j)^{-r_{ij}}, \quad \sum_{i=1}^m \theta_i = 1, \quad (3)$$

where  $\theta_i > 0 (i = 1, 2, \dots, m)$  are the treatment parameters.

### 3. The Choice of an Informative Prior Distribution

The appropriate model of paired comparison having been selected and the likelihood thereby determined, it is necessary to choose a prior distribution for the parameters of the model. Typically this will be supposed to belong to a parametric family; when the problem reduced to assessing these parameters. To distinguish them from the parameters in the likelihood, they are termed hyperparameters. We consider the Bradley-Terry model having a likelihood (3). The prior is supposed to belong to the member of the Dirichlet family with density:

$$p(\boldsymbol{\theta}) = B(a_1, a_2, \dots, a_m)^{-1} \prod_{i=1}^m \theta_i^{a_i-1}, \quad (4)$$

where  $\boldsymbol{\theta} = (\theta_1, \theta_2, \dots, \theta_m)$  are treatment parameters,  $\sum_{i=1}^m \theta_i = 1$ ,  $B(a_1, a_2, \dots, a_m)$  stands for a generalization of the beta function and  $a_i (i = 1, 2, \dots, m)$  are the hyperparameters.

### 4. The Choice of a Non-informative Prior Distribution

When prior elicitation is difficult or little prior information is available, one may consider analysis with conventionally chosen priors that may reflect little prior information. These priors are known as non-informative priors or indifference or vague or reference priors. Berger (1985) argues that Bayesian analysis using non-informative priors is the single most powerful method of statistical analysis, in the sense of being the adhoc method most likely to yield a sensible answer. This topic has an extensive literature, eg. Jeffreys (1946), Bernardo (1979), Ghosh and Mukerjee (1992), Kass (1990), Clarke and Wasserman (1993) and Aslam (1995).

#### 4.1 The Non-informative Prior from 'Jeffreys Rule'

This is defined as the density of the parameters proportional to the square root of the determinant of the Fisher information matrix. Symbolically, let  $\boldsymbol{\phi} = (\phi_1, \dots, \phi_m)^t$  is a vector of parameters, the Jeffreys prior distribution  $\{p_j(\boldsymbol{\phi})\}$  is defined as:

$$p_j(\boldsymbol{\phi}) \propto \sqrt{\det\{I(\boldsymbol{\phi})\}}, \quad (5)$$

where 'det' denotes the determinant and  $I(\boldsymbol{\phi})$  is the  $(m \times m)$  Fisher information matrix which is obtained as:

$$I_{ij}(\boldsymbol{\phi}) = -E \left\{ \frac{\partial^2 \ln L(\boldsymbol{\phi})}{\partial \phi_i \partial \phi_j} \right\}, \quad (6)$$

where E denotes expectation on data and  $i$  and  $j$  stand for rows and columns of determinant, respectively.

## 4.2 Properties of the Non-informative Jeffreys Prior

The Jeffreys prior shows many nice properties that make it an attractive non-informative prior. The Jeffreys prior has an invariance property with respect to a one-to-one transformation of the parameters in the sense that we get consistent answer in any parameterization. The Jeffreys prior works well in one-dimensional problem but needs adhoc modification for multi-dimensional case as suggested by Jeffreys (1946) himself. Bernardo (1979) shows that Jeffreys prior is an appropriate reference prior if there are no parameters which are regarded as nuisance parameters and if the (joint) posterior distribution of all the parameters is asymptotically normal. Kass (1990) explains that a main feature of the Jeffreys prior is that it is a uniform measure in an information metric which can be regarded as the natural metric for statistical inference. Another important aspect of the Jeffreys prior is that it is not affected by a restriction on the parameter space. Some of the applications of the Jeffreys prior are given in Ibrahim and Laud (1991) and Poirier (1994).

## 4.3 The Non-informative Jeffreys Prior for the Parameters of the Model

The appropriateness of the Jeffreys prior requires that there should be no parameters which are regarded as nuisance parameters and the posterior distribution of all the parameters should be asymptotically normal. The likelihood function (3) of the model can be represented in the following form:

$$l(\mathbf{x}; \theta_1, \dots, \theta_m) = \exp \left\{ \sum_{i=1}^m n_i \ln \theta_i - \sum_{i(<j)=1}^m r_{ij} \ln(\theta_i + \theta_j) \right\}. \quad (7)$$

The likelihood function (7) belongs to the exponential family and it follows from the regularity conditions of Johnson and Ladalla (1991) that the posterior distribution is asymptotically normal. Hence, here is no parameter which is regarded as 'nuisance' so the Jeffreys prior is the appropriate choice of non-informative prior and hereafter will called NJP ((NJP stands for non-informative Jeffreys prior) distribution.

The NJPs of the parameters for the case of two, three and four treatments in the model are derived. Some transformations and techniques are applied to understand and to find a simple symmetric closed form for the case of three and four treatments.

Case (i) *When the number of treatments  $m=2$ :*

In this case, the treatment parameters are  $\theta_1$  and  $\theta_2$ . Consider  $\theta_1 = \theta$  and  $\theta_2 = 1 - \theta$ . Now the likelihood function for the observed outcomes  $\mathbf{x}$  from (3) is:

$$l(\mathbf{x}; \theta) = \frac{r_{12}! \theta^{n_{12}} (1-\theta)^{n_{21}}}{n_{12}! n_{21}!}, \quad (8)$$

which is the binomial model:  $n_{12} \sim \text{Bin}(r_{12}, \theta)$ .

The Fisher's information for the likelihood function (8) is obtained as:

$$I(\theta) = -E \left\{ \frac{\partial^2 \ln l(\mathbf{x}; \theta)}{\partial \theta^2} \right\} = \frac{1}{\theta(1-\theta)}. \quad (9)$$

The NJP  $\{ p_J(\theta) \}$  for the parameter  $\theta$  of the model is:

$$p_J(\theta) = \frac{1}{\pi\sqrt{\theta(1-\theta)}}, \quad 0 \leq \theta \leq 1. \tag{10}$$

The graph of the NJP (10) is presented in Figure 1.

Case (ii). When the number of treatments  $m=3$ :

Here the treatment parameters are  $\theta_1, \theta_2$  and  $\theta_3$ . Using the constraint  $\theta_3 = 1 - \theta_1 - \theta_2$ , the likelihood is:

$$l(\mathbf{x}; \theta_1, \theta_2) = \binom{r_{12}}{n_{12}} \binom{r_{13}}{n_{13}} \binom{r_{23}}{n_{23}} \frac{\theta_1^{n_{12}+n_{13}} \theta_2^{n_{21}+n_{23}} (1-\theta_1-\theta_2)^{n_{31}+n_{32}}}{(\theta_1+\theta_2)^{r_{12}} (1-\theta_1)^{r_{23}} (1-\theta_2)^{r_{13}}}. \tag{11}$$

A program is designed in the MAPLE package to get the NJP for an unbalanced design that the numbers of comparisons for each pair are not equal. We obtain the following form for the NJP:

$$p_J(\theta_1, \theta_2) \propto \frac{g(\theta_1, \theta_2)}{(\theta_1 + \theta_2)(1 - \theta_1)(1 - \theta_2)\sqrt{\theta_1\theta_2(1 - \theta_1 - \theta_2)}}, \tag{12}$$

$$\theta_1, \theta_2 \geq 0, \theta_1 + \theta_2 < 1,$$

where  $g(\theta_1, \theta_2) = [r_{13}\theta_1^3(r_{12} - r_{23}) + r_{23}\theta_2^3(r_{12} - r_{13}) + r_{13}\theta_1^2(r_{23} - 2r_{12}) + r_{23}\theta_2^2(r_{13} - 2r_{12}) + r_{12}(r_{13}\theta_1 + r_{23}\theta_2) - r_{23}\theta_1\theta_2\{3r_{12}(\theta_1 + \theta_2) + 2r_{13}\}]^{\frac{1}{2}}$ .

If a balanced design is considered i.e.  $r_{ij}=r$  ( $i < j=1,2,3$ ), then 'r' can be merged in a normalizing constant, so the NJP is:

$$p_J(\theta_1, \theta_2) = \frac{C\{(\theta_1 + \theta_2)(1 - 3\theta_1\theta_2) - (\theta_1 - \theta_2)^2\}^{\frac{1}{2}}}{(\theta_1 + \theta_2)(1 - \theta_1)(1 - \theta_2)\{\theta_1\theta_2(1 - \theta_1 - \theta_2)\}^{\frac{1}{2}}}, \tag{13}$$

$$\theta_1, \theta_2 \geq 0, \theta_1 + \theta_2 < 1,$$

where 'C=0.052798' is evaluated by numerical integration.

It is noted that the parameters  $\theta_1$  and  $\theta_2$  are exchangeable which is clear algebraically in (13).

The closed forms for the marginal NJPs of  $\theta_1$  and  $\theta_2$  are intractable and we evaluate the marginal NJPs of  $\theta_1$  and  $\theta_2$  numerically. Thus the marginal NJP of  $\theta_1$  is:

$$p_J(\theta_1) = C \int_0^{1-\theta_1} \frac{\{(\theta_1 + \theta_2)(1 - 3\theta_1\theta_2) - (\theta_1 - \theta_2)^2\}^{\frac{1}{2}} d\theta_2}{(\theta_1 + \theta_2)(1 - \theta_1)(1 - \theta_2)\{\theta_1\theta_2(1 - \theta_1 - \theta_2)\}^{\frac{1}{2}}}, \quad 0 \leq \theta_1 \leq 1, \tag{14}$$

Similarly, the marginal Jeffreys prior of  $\theta_2$  can be obtained. The graph of the marginal NJP of  $\theta_1$  is given in Figure 2.

We have found the marginal density of the NJP out of curiosity and to see whether we can believe, it is appropriate as a consequence of the NJP. In terms of calculating our 'answer' to the overall analysis of the data, the calculation of these marginal is not necessary. Some transformations and techniques are used to find out a simple closed form for the marginal NJP of  $\theta_1$  (see Appendix 1). We conclude that although the graph of the marginal density of  $\theta_1$  has a simple appearance, it does not appear to be expressible in a closed form or have any reasonably elegant or simple characterization.

Case (iii) *When the number of treatments  $m=4$ :*

The likelihood for the case of four treatments can be written from (3). The treatment parameters are  $\theta_1, \theta_2, \theta_3$  and  $\theta_4$ , the NJP for the parameters  $\theta_1, \theta_2$  and  $\theta_3$  is derived with the identifiability constraint:  $\theta_1 + \theta_2 + \theta_3 + \theta_4 = 1$ . A program similar to the case of 3 treatments is written in the MAPLE package to obtain the NJP for a balanced design. The derived NJP (given in Appendix 3) has a very complicated form and does not seem to be easy for application.

However we can determine the NJP numerically via numerical partial differentiation of the log likelihood function. A program for computing NJP numerically is written in the SAS package which can be seen in Aslam (1995).

The graph of the marginal NJP (without a constant of proportionality) for the parameter  $\theta_1$ , is given in Figure 3 which is obtained through the program in the SAS package.

**NJP FOR THE PARAMETER  $\theta$  OF THE MODEL WHEN  $M=2$**

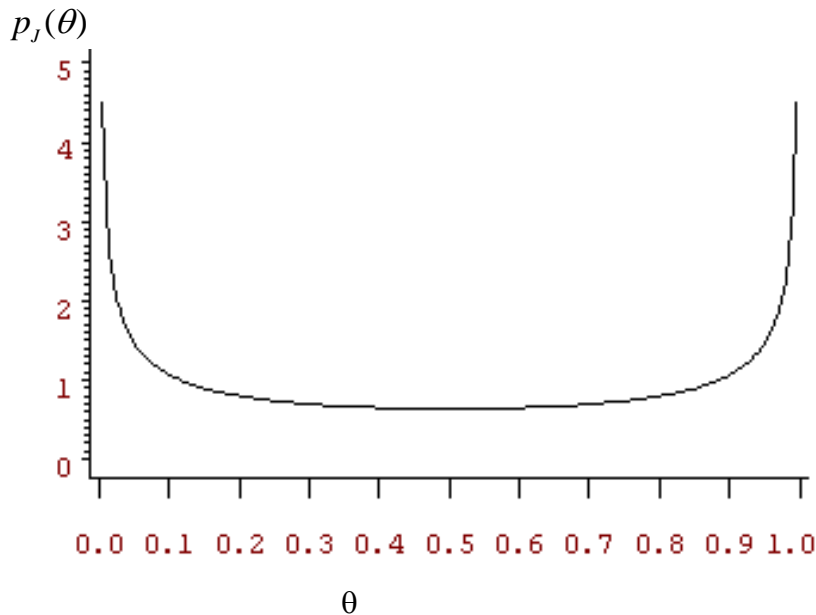


Figure 1

**MARGINAL NJP FOR  $\theta_1$  OF THE MODEL WHEN  $M=3$**

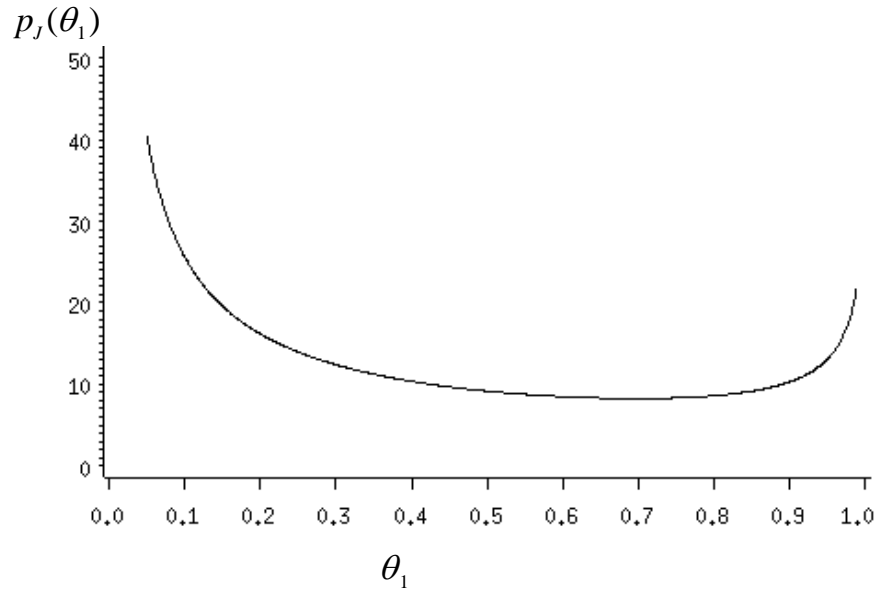


Figure 2

**MARGINAL NJP FOR  $\theta_1$  OF THE MODEL WHEN  $M=4$**

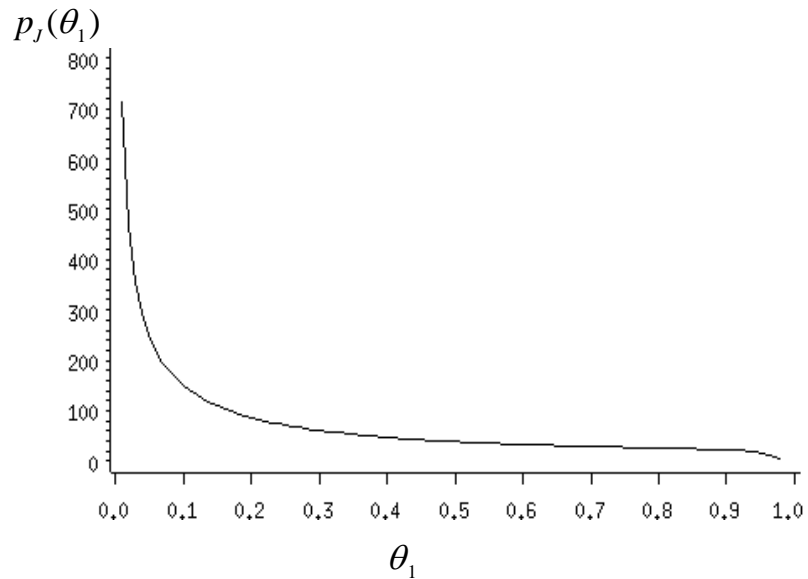


Figure 3

**5. Comparison of the NJP and the Dirichlet Prior**

The NJP and the informative prior Dirichlet distribution are compared through the Lindley-Shannon information (LSI) {Lindley (1956) and Shannon (1948)} (see Appendix 2 for definition of LSI). Our choice of informative prior for the parameters belongs to the Dirichlet family, so for comparison with NJP, we choose the Dirichlet distribution. The LSI of the Dirichlet distribution is computed for the values of hyperparameters from 0 to

5. We consider this range because the range of Lindley-Shannon information generated includes the LSI of the NJP in each case ( $m=2, 3, 4$ ). The LSI is used because it is a fundamental measure of information (with an axiomatic basis of justification for its choice).

(a) *The number of treatments  $m=2$*

The LSI for the NJP (10) has the following form:

$$I_J = \pi^{-1} \int_0^1 \{\theta(1-\theta)\}^{-1/2} \ln[\pi^{-1}\{\theta(1-\theta)\}^{-1/2}] d\theta. \quad (15)$$

The closed form for (14) is intractable and convergence is very slow with numerical integration. We use the transformation:  $\theta = (1 + e^{-z})^{-1}$  for fast convergence then the LSI is:

$$I_J = \frac{1}{\pi} \int_{-\infty}^{+\infty} \frac{e^{z/2}}{(1 + e^z)} \ln\left(\frac{1 + e^z}{\pi e^{z/2}}\right) dz, \quad (16)$$

A closed form for the integral (16) is also intractable but numerically  $I_J = 0.24156$ . The LSI for the informative prior beta distribution  $B(\alpha, \beta)$  with equal hyperparameters ( $\alpha = \beta = c$ ) from '0.1' to '5' is given in the Table given below. The LSI of the NJP corresponds to  $c=0.5$  in  $B(c, c)$ . We note that although the NJP satisfies an invariance property a uniform distribution  $B(1,1)$  has a small LSI. Generally, we can conclude that a  $B(c, c)$  prior with 'c' between 0.5 and 1.5 represents very little prior information. This indicates that the NJP is a reasonable choice from the standpoint of the LSI and is compatible with other 'similar' contenders.

(b) *The number of treatments  $m=3$*

The LSI of the NJP (13) is compared with the LSI of the prior Dirichlet distribution  $D(a_1, a_2, a_3)$  for equal hyperparameters ( $a_1 = a_2 = a_3 = c$ ). It is found that the LSI is equal to 1.46817 for the NJP whereas the LSI for the prior Dirichlet distribution is given in the Table given below. Comparison of the LSI for a  $D(c, c, c)$  distribution with the value 1.46817 of the NJP indicates that the Dirichlet distribution is a reasonable choice when there is little prior information for the values of 'c' between 0.5 and 4.

(c) *The number of treatments  $m=4$*

Now consider the Dirichlet distribution  $D(c, c, c, c)$  of equal hyperparameters for the case of  $m=4$ . The LSI for the values of 'c' from 0.1 to 5 is given in the Table. The LSI of the NJP is obtained to be 2.53346. This shows that the Dirichlet Distribution  $D(c, c, c, c)$  is a sensible choice for a non-informative prior when  $c < 2.5$ .



**LINDLEY-SHANNON INFORMATION  
for the Beta, Dirichlet Distributions and NJPs**

$c$	$m=2$ $B(c, c)$	$m=3$ $D(c, c, c)$	$m=4$ $D(c, c, c, c)$
0.1	4.59343	8.27150	0.26069
0.2	2.04563	4.53243	1.04444
0.3	0.95036	2.52865	1.59304
0.4	0.47422	1.61334	1.82118
0.5	0.24156	1.16334	1.87040
0.6	0.11942	0.92494	1.85005
0.7	0.05384	0.79766	1.81554
0.8	0.01971	0.73143	1.78927
0.9	0.00415	0.70120	1.77780
1.0	0.00000	0.69315	1.78103
1.2	0.01083	0.71422	1.82174
1.4	0.03438	0.76013	1.89110
1.6	0.06324	0.81646	1.97452
1.8	0.09402	0.87664	2.06360
2.0	0.12509	0.93749	2.15379
2.5	0.19981	1.08411	2.37169
3.0	0.26786	1.21802	2.57070
4.0	0.38450	1.44813	2.91349
5.0	0.48064	1.63831	3.19717
NJP	0.24156	1.46817	2.53346

**6. Conclusion**

For 2 parameters case, it is concluded that a prior Beta distribution  $B(c, c)$  with hyperparameter 'c' between 0.5 and 1.5 represents very little prior information. This indicates that the NJP is a reasonable choice from the standpoint of the LSI and is compatible with other 'similar' contenders. When 3 parameters are considered then the LSI for the prior Dirichlet distribution  $D(c, c, c)$  indicates that the Dirichlet distribution is a reasonable choice for the values of 'c' between 0.5 and 4. For 4 parameters case, this shows that the prior Dirichlet Distribution  $D(c, c, c, c)$  is a sensible choice when  $c < 2.5$ .

## References

1. Aslam, M. (1995). Bayesian analysis for paired comparison data, Ph. D. thesis, Aberystwyth: University of Wales. U. K.
2. Abbas, N. (2010). Paired comparison modeling with Bayesian analysis, Ph. D. thesis, Quaid-i-Azam University, Pakistan.
3. Altaf, S., Aslam, M. and Aslam, M. (2011). Bayesian analysis of the Davidson model for paired comparison with order effect using noninformative priors. *Pak. J. of Statistics*, 27(2), 171-185.
4. Berger, J. O. (1985). *Statistical decision theory and Bayesian analysis*, 2nd edn. New York: Springer-Verlag.
5. Berger, J. O. and Bernardo, J. M. (1989). Estimating a product of means: Bayesian analysis with reference priors, *J. Amer. Statist. Assoc.* 84, 200-207.
6. Bernardo, J. M. (1979). Reference posterior distribution for Bayesian inference. *J. Roy. Statist. Soc., Ser. B*, 41, 113-147.
7. Bradley, R. A. and Terry, M. B. (1952). Rank analysis of incomplete block design: I, the method of paired comparisons, *Biometrika*, 39, 324-345.
8. Box, G. E. P. and Tio, G. C. (1973). *Bayesian Inference in statistical analysis*, Reading, Massachusetts: Addison-Wesley.
9. Clarke, B. and Wasserman, L. (1993). Noninformative prior and nuisance parameters. *J. Amer. Statist. Assoc.* 88, 1427-1434.
10. David, H. A. (1988). The method of paired comparisons, 2nd edn., Charles Griffin and Company-London.
11. Gilani, G. M. and Abbas, N. (2008). Bayesian analysis for the paired comparison model with order effects (using NIP). *Pak. J. Statis. and Oper.Res.*, 4(2), 83-91.
12. Ghosh, J. K. and Mukerjee, R. (1992). Noninformative priors. In: *Bayesian Statistics 4*, Eds. J. M. Bernardo, J. O. Berger, A. P. Dawid, and A. M. F. Smith, Oxford: Oxford University Press. 195-210.
13. Ibrahim, J. G. and Laud, P. W. (1991). On Bayesian analysis of general linear model using Jeffreys' prior, *J. Amer. Statist. Assoc.* 86, 081-986.
14. Jeffreys, H. (1946). An invariant form for the prior probability in estimation problems. *Proceedings of the Roy. Soc. of London*, Ser. A, 186, 453-461.
15. Johnson, R. A. and Ladalla, J. N. (1979). The large sample behaviour of posterior distributions when sampling from multiparameters exponential family models and allied results, *Sankhya*, B, 41, 196-215.
16. Kass, R. E. (1990). Data-translated likelihood and Jeffreys' rule, *Biometrika*, 77, 17-114.
17. Lindley, D. V. (1956). On a measure of the information provided by an experiment, *The Annals of Mathematical Statistics*, 27, 986-1005.
18. Poirier, D. (1994). Jeffreys prior for logit models. *Journal of Econometrics*, 63, 327-339.
19. Shannon, C. E. (1948). A mathematical theory of communication, *Bell System Technical Journal*, 27, 379-423 and 623-656.

## Appendix

1. (a) In order to try and characterize, the marginal distribution of  $\theta_1$  (14), we consider the implied distribution of  $\phi_i$  (i-1, 2, 3) where  $\phi_i$  is one of the following functions:

$$(i) \phi_1 = \ln \theta_1, (ii) \phi_2 = \ln \left\{ \frac{\theta_1}{(1-\theta_1)} \right\}, \quad (iii) \phi_3 = \Phi^{-1}(\theta_1),$$

where  $\Phi$  is the cumulative distribution function of the standard normal distribution. It is observed that the graphs of the functions (ii) and (iii) are approximately normally distributed.

(b) The marginal NJP  $p_j(\theta_1)$  is compared with the beta distribution. It is observed that the marginal Jeffreys prior of  $\theta_1$  approximately follows the beta distribution.

(c) The graphs of logarithm and reciprocal of the marginal NJP of  $\theta_1$  are drawn but these haven't standard (well-known) functional shape.

## 2. Lindley-Shannon Information

Shannon (1948) has proposed the theory of information in the context of communication engineering. He introduces the idea that information is a statistical concept. Now measures of information are closely related to ideas of uncertainty and probability. Lindley (1956) suggests the use of information as a statistical criterion in the design of experiments. He explains that one of the purposes of experimentation is to gain knowledge about the state of nature (that is, about the parameters) without having specific actions in mind. The knowledge is measured by the amount of information.

Let  $p(\theta)$  be a prior distribution, the amount of information  $I\{p(\theta)\}$  is defined as:

$$I\{p(\theta)\} = \int_{\theta} p(\theta) \ln\{p(\theta)\} d\theta, \quad (2.1)$$

whenever the integral exists. For any value of  $\theta$  for which  $p(\theta)=0$ , define  $p(\theta) \ln\{p(\theta)\}$  to be zero [appropriate since  $\lim_{x \downarrow 0} \{x \ln(x)\} = 0$ ].

An alternative but equivalent form of (2.1) is

$$I\{p(\theta)\} = E_{\theta} \ln\{p(\theta)\}, \quad (2.2)$$

where  $E_{\theta}$  denotes the expectation operator with respect to  $\theta$ .

### 3. NJP of the Parameters for the Model When $m=4$

Consider,  $t_i = \theta_i$ ;  $i=1,2,3$ , the Jeffreys prior obtained by MAPLE package is:

$$\begin{aligned}
 \text{jp4} := & \left( (-8 t_1^2 t_2 t_3 + t_1^4 t_3^2 t_2 + 40 t_1^3 t_3^2 t_2^2 + 11 t_1^4 t_3^3 t_2^3 \right. \\
 & - 44 t_1^3 t_3^3 t_2^2 + 40 t_2^3 t_3^2 t_1^2 - 44 t_2^3 t_3^3 t_1^3 + 11 t_2^4 t_3^3 t_1^4 \\
 & - 18 t_2^2 t_3^4 t_1^2 + 11 t_2^3 t_3^4 t_1^3 + t_1^2 t_3^4 t_2^4 + 3 t_1^2 t_3^5 t_2^5 \\
 & + 3 t_1^5 t_3^2 t_2^2 + 11 t_1^3 t_3^4 t_2^3 - 12 t_1^3 t_3^3 t_2^3 + 11 t_1^4 t_3^4 t_2^3 \\
 & + 11 t_1^3 t_3^4 t_2^4 - 18 t_1^4 t_3^2 t_2^2 t_3 - 44 t_1^3 t_3^3 t_2^4 + t_1^4 t_3^2 t_2^3 \\
 & + 3 t_1^5 t_3^2 t_2^3 + 3 t_1^5 t_3^2 t_2^3 + 12 t_1^5 t_3^2 t_2^3 - 2 t_1^6 t_3^2 t_2^3 \\
 & - 2 t_1^6 t_3^2 t_2^3 + 24 t_1^3 t_3^3 t_2^3 + 3 t_1^6 t_3^2 t_2^3 + 3 t_1^6 t_3^2 t_2^3 + 3 t_1^2 t_3^6 \\
 & + t_2^4 t_3^2 t_1^2 + t_2^2 t_3^4 t_1^4 + 12 t_1^5 t_3^2 t_2^3 + t_2^5 t_3^4 t_1^2 + 12 t_2^5 t_3^5 t_1^5 \\
 & + 3 t_2^5 t_3^2 t_1^2 - 2 t_2^6 t_3^3 t_1^3 - 18 t_2^4 t_3^2 t_1^2 + 3 t_2^5 t_3^2 t_1^2 + t_1^4 t_3^4 \\
 & + t_2^6 t_3^3 - 6 t_2^5 t_3^2 t_3 + 3 t_2^6 t_3^2 t_3 - 6 t_1^5 t_3^2 t_2^2 - 16 t_1^4 t_3^3 t_2^3 - 16 t_1^3 t_3^4 t_2^4 \\
 & + 9 t_1^5 t_3^3 t_2^3 + 12 t_1^4 t_3^4 t_2^4 + 9 t_1^3 t_3^5 t_2^5 - 6 t_1^5 t_3^2 t_2^4 - 16 t_1^4 t_3^3 t_2^3 \\
 & + 9 t_1^5 t_3^3 t_2^3 - 6 t_1^2 t_3^5 t_2^5 - 16 t_2^4 t_3^3 t_1^3 + 9 t_2^5 t_3^3 t_1^3 - 16 t_2^3 t_3^4 t_1^4 \\
 & + 12 t_2^4 t_3^4 t_1^4 - 16 t_1^3 t_3^4 t_2^3 - 6 t_1^2 t_3^5 t_2^5 + 9 t_1^3 t_3^5 t_2^5 - 6 t_2^2 t_3^5 \\
 & + 9 t_2^3 t_3^5 t_2^3 + 12 t_1^4 t_3^4 t_2^3 - t_2^3 t_3^3 - t_1^3 t_2^3 - t_1^3 t_3^3 - 2 t_1^2 t_2^2 \\
 & \left. - 2 t_2^2 t_3^2 - t_1^3 t_3^3 - t_2^3 t_3^3 - 2 t_1^2 t_3^2 + 3 t_2^4 t_3^4 - 3 t_2^5 t_3^5 \right)
 \end{aligned}$$

*Lindley-Shannon Information for Comparison of Priors Under Paired Comparisons Model*

$$\begin{aligned}
 & + t_2^4 t_3^2 + 4 t_1^3 t_2^2 + 4 t_1^3 t_2^3 + 4 t_1^3 t_3^2 + 4 t_1^3 t_3^3 + 4 t_1^2 t_2^3 \\
 & + t_1^2 t_2^4 + 4 t_1^3 t_2^2 t_3 + 4 t_1^2 t_2^3 t_3 + 4 t_1^3 t_2^3 t_3 + t_1^2 t_2^4 t_3 + 3 t_1^4 t_2^3 \\
 & - 3 t_1^5 t_2^3 + 3 t_1^4 t_2^4 t_3 - 3 t_1^5 t_2^3 t_3 + 4 t_1^2 t_1^3 t_3 + t_1^2 t_1^4 t_3 - t_1^3 t_2^3 \\
 & + 3 t_1^4 t_2^5 t_1 - 3 t_1^5 t_2^6 t_1 + t_1^6 t_1 + 3 t_1^6 t_2^2 t_1 + 3 t_1^4 t_1^2 t_2 - 3 t_1^5 t_1^2 t_2 \\
 & + t_1^4 t_1^2 t_2 + 3 t_1^4 t_1^4 t_3 - 3 t_1^5 t_1^5 t_3 - 8 t_1^2 t_1^2 t_2^3 - 8 t_1^2 t_1^2 t_2^3 \\
 & - 22 t_1^2 t_1^3 t_2^3 - 46 t_1^2 t_1^2 t_2^2 t_3 - 22 t_1^2 t_1^3 t_2^3 t_3 - 22 t_1^3 t_1^2 t_2^3 t_3 \\
 & - 22 t_1^3 t_1^2 t_2^3 t_3 - 22 t_1^3 t_1^2 t_3^2 t_2 - 26 t_1^4 t_1^3 t_2^3 + 24 t_1^3 t_1^2 t_2^3 t_3 \\
 & + 26 t_1^2 t_1^2 t_2^3 t_3 - 26 t_1^4 t_1^2 t_2^3 t_3 - 22 t_1^3 t_1^2 t_2^3 t_3 + 24 t_1^3 t_1^2 t_2^3 t_3 \\
 & + t_1^6 t_1^3 t_3 + t_1^6 t_2^3 t_3 + t_1^6 t_1^2 t_2 + t_1^6 t_1^3 t_3 + 3 t_1^2 t_1^6 t_3 + 26 t_1^2 t_1^2 t_3 \\
 & + 26 t_1^2 t_1^2 t_2^3 t_3) / (t_1^2 (t_2 + t_3)^2 t_3 t_2 (-1 + t_1 + t_2 + t_3) \\
 & (-1 + t_1 + t_2)^2 (-1 + t_1 + t_3)^2 (-1 + t_2 + t_3)^2 (t_1 + t_3)^2 \\
 & (t_1 + t_2)^2)^{1/2}
 \end{aligned}$$