Bayesian modeling to paired comparison data via the Pareto distribution

Abstract

A probabilistic approach to build models for paired comparison experiments based on the comparison of two Pareto variables is considered. Analysis of the proposed model is carried out in classical as well as Bayesian frameworks. Informative and uninformative priors are employed to accommodate the prior information. Simulation study is conducted to assess the suitablily and performance of the model under theoretical conditions. Appropriateness of fit of the is also carried out. Entire inferential procedure is illustrated by comparing certain cricket teams using real dataset.

Key words: Pair-wise comparisons, Pareto distribution, Bayesian Analysis, Worth parameter, ML estimates.

Mathematics Subject Classification: 62J15, 62F07; 62F10; 62E15.

1. Introduction

The method of paired comparisons (PC) is a technique in which items (objects, stimulus, treatments, options, etc) are presented to one or more judges (respondents, raters, jurists, etc) in pairs. The judges may prefer one of the two items or declare a tie. The PC method is primarily used for subjective judgments where quantitative measurement is impossible or impracticable. Hence it is widely used by psychometricians. The most frequent application has been to sensory testing; especially taste testing, consumer tests, personal rating in sports, choice behaviors, etc. If several items are to be compared simultaneously, then it may be done preferably by simple ranking. But if the differences between items are small, the PC method is the best choice to avoid extraneous source of variation caused by a third object and hence a finer decision may be reached.

The PC technique can be useful for ranking items related to particular issue that are too numerous or too similar to rank mentally. David (1988) provides a detailed review on the PC method. Probably, the most-cited among the applied uses of PC is the tournament analysis in which items are players or teams competing with each other in pairs.

Different PC models are based on the mechanism of production of sensations/responses from human brain regarding the items being compared. Thurstone (1927) assumes responses to follow normal distribution but Bradley and Terry (1952) consider the Logistic distribution in proposing their models. Thompson and Singh (1967) arrive at the Thurstone-Mosteller model by regarding the registered merit or experienced sensation as an average or median of a large number of signals transmitted to judges' brain. Stern (1990) builds his models in terms of probability that one gamma random variable with a fixed shape parameter is less than another independent gamma random variable with the same shape parameter but a different scale parameter. Different values of the shape parameters provide different PC models. Rao and Kupper (1967), Davidson (1970) and Glenn and David (1960) extend the basic models by including ties. Henery (1981, 1986) and Mak (1985) deal with the order statistics models.

Abbas and Aslam (2011a) extend the renowned Bradley and Terry (1952) model by accommodating quantitative weights in the qualitative paired comparisons. Abbas and Aslam (2012a) model the factors influencing the judges' evaluations of items through mixture models for the preference datasets. The probabilistic features of the mixture distribution are addressed and inferential and computational issues emerging out of the maximum likelihood estimation are dealt with. A methodology is proposed by Abbas et al. (2011b) to measure the actual comparative worth of the competing items on a finer scale by assigning some refined ranks to each of the two competing items on a finer scale. The assigned ranks are then converted to a refined paired comparison dataset in the form of preference matrix to be used for ranking the items.

The break-up of the study is as follows: In Section 2, the logical basis and ideas to construct the Pareto model are discussed. It also compares our proposed model with those developed by Stern (1990). Section 3 deals with likelihood function and notations used in this study. The classical maximum likelihood (ML) estimates along with their standard errors are presented in Section 4. Section 6 pertains to uninformative and informative priors used in deriving the posterior distribution. The elicitation of hyperparameters – the parameters of priors – is also conducted therein. The posterior distributions and the marginal posterior distributions are studied in Section 7. Section 8 is concerned with the Bayesian analysis of the proposed model. It comprises estimating the posterior means, preference probabilities, posterior probabilities of the hypothesis and predictive probabilities of preference for future comparisons. Section 9 is reserved for the simulation study of the proposed model to assess the performance. Section 10 compares the priors used in the Bayesian analysis. The appropriateness of the model is viewed in Section 11. Section 12 concludes and discusses the entire study.

2. The Pareto model

The Pareto distribution, named after the Italian economist Vilfredo Pareto, is a power law probability distribution found in large number of real world situations and was originally used by him to describe allocation of wealth among different individuals. Obviously, a large portion of the wealth of a society is owned by a smaller percentage of people of a society. DeGroot (1970) also declares that the Pareto distribution can be used as an income model. Literature also furnishes with the situations where the distribution can be used to model words lengths in larger paragraphs, the value of oil reserves in oil fields, size of sand particles, standardized price returns on individual stocks, etc, [Logan (2012)].

The motivation of using the Pareto distribution is based on the fact that it models income distribution and a Pareto variable has its minimum value equal to its location parameter. So the locality with a greater minimum income (location parameter) may be preferred to that with a lower minimum income, while both having same shape parameter. It is in accordance with the model building pattern adopted by Stern (1990), who compares the waiting times of two point-scoring competitions using gamma distributions with same shape parameters and different location parameters and the player taking less time to score points is declared a winner.

The Pareto distribution with scale parameter α and location parameters θ_i (specifying the least values of the Pareto variable X_i) is given by

$$f\left(x_i\right) = \alpha\theta_i^{\alpha} \big/ x_i^{\alpha+1}, \alpha > 0, x_i > \theta_i > 0 .$$

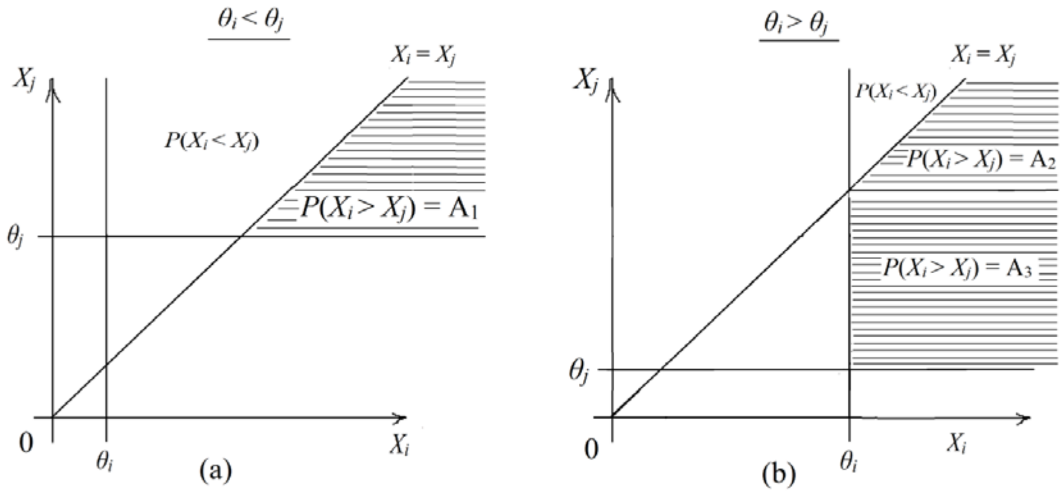

Figure 1: Derivation of the model

Following the model developing criterion adopted in Stern (1990), the scale parameter α is considered common for both of the competitors and the location parameters vary. For $\theta_i \le \theta_j$, the probability of preferring treatment 'i' over 'j ', denoted by ϕ_{ij}, is

$$\phi_{ij} = P(X_i > X_j) = A_1 = \int_{\theta_j}^{\infty}\int_{x_j}^{\infty}\left(\frac{\alpha\theta_j^{\alpha}}{x_j^{\alpha+1}}\right)\left(\frac{\alpha\theta_i^{\alpha}}{x_i^{\alpha+1}}\right)dx_i dx_j = \frac{1}{2}\left(\frac{\theta_i}{\theta_j}\right)^{\alpha}, \tag{2.1}$$

But when $\theta_i > \theta_j$, it becomes

$$\phi'_{ij} = P(X_i > X_j) = A_2 + A_3$$

$$= \int_{\theta_i}^{\infty}\int_{x_j}^{\infty}\left(\frac{\alpha\theta_j^{\alpha}}{x_j^{\alpha+1}}\right)\left(\frac{\alpha\theta_i^{\alpha}}{x_i^{\alpha+1}}\right)dx_i dx_j + \int_{\theta_j}^{\theta_i}\int_{\theta_i}^{\infty}\left(\frac{\alpha\theta_j^{\alpha}}{x_j^{\alpha+1}}\right)\left(\frac{\alpha\theta_i^{\alpha}}{x_i^{\alpha+1}}\right)dx_i dx_j = 1 - \frac{1}{2}\left(\frac{\theta_j}{\theta_i}\right)^{\alpha}, \tag{2.2}$$

where θ_i and θ_j denote the location (worth) parameters for the competing items. Here $\phi_{ji} = 1 - \phi_{ij}$ or $\phi'_{ji} = 1 - \phi'_{ij}$ denote the probability of preferring item j over i (ties not being allowed). Expressions (2.1) and (2.2) serve as a full-blown PC model. Since, different Pareto models are obtained assuming different values of the parameter α, so it becomes necessary to see how different values of α affect the preference probabilities ϕ_{ij} and the worth parameters θ_i, for $i = 1,...,t$.

Now it is vital to elicit a value for the parameter α that best suits a certain data set. To accomplish this, we have used the minimum chi-square method suggested by Abbas and Aslam (2013). It is inferred therein that α =0.55 produced minimum chi-square value for the given data set and is hence used in the subsequent analyses.

3. Likelihood function

Let $n_{ij} = n_{ji}$ be the total number of comparisons made between item *i* and *j*, a_{ij} be the number of times item *i* is preferred to *j* and $a_{ji} = n_{ij} - a_{ij}$ denotes its reverse. For the present situation, the trials are independent with only two categories of the outcomes for all trials, e.g., preferring item *i* over *j*, or the vice versa, and each outcome has a constant probability ϕ_{ij} of preferring item *i* over *j* for all $i(\neq j) = 1, 2, ..., t$. The PC experiment is performed a fixed number of times n_{ij}. So the variable a_{ij} follows a Binomial distribution $B(a_{ij}; n_{ij}, \phi_{ij})$ and the likelihood function $L(\boldsymbol{a}; \boldsymbol{\theta})$ is:

$$L(\mathbf{a}\boldsymbol{\theta}) = \underbrace{\prod_{i(<j)=i}^{h} C_{a'_{ij}}^{n_{ij}} \phi'^{a'_{ij}}_{ij} \left(1-\phi'_{ij}\right)^{n_{ij}-a'_{ij}}}_{\text{For } \theta_i > \theta_j} \underbrace{\prod_{i(<j)=1}^{t} C_{a_{ij}}^{n_{ij}} \phi_{ij}^{a_{ij}} \left(1-\phi_{ij}\right)^{n_{ij}-a_{ij}}}_{\text{For } \theta_i < \theta_j} \quad (3.1)$$

4. ML estimation

The parametric ML estimates of the proposed model may be found by maximizing the likelihood function (3.1) with regards to the unknown model parameters. As the logarithm is a non-decreasing function, hence any function and its logarithm are maximized at the same points. However, the maximization of the logarithm of a function is analytically easier than the direct maximization. So instead of maximizing (3.1), we maximize

$$l = \sum_{i(<j)=1}^{h} \left[\log C_{a'_{ij}}^{n_{ij}} + a'_{ij} \log \phi'_{ij} + (n_{ij} - a'_{ij}) \log \phi'_{ji} \right] + \sum_{i(<j)=h+1}^{t} \left[\log C_{a_{ij}}^{n_{ij}} + a_{ij} \log \phi_{ij} + (n_{ij} - a_{ij}) \log \phi_{ji} \right] \quad (4.1)$$

$$l = \sum_{i(<j)=1}^{h} \left[\log C_{a'_{ij}}^{n_{ij}} + a'_{ij} \log\{1 - \tfrac{1}{2}(\theta_j/\theta_i)^{\alpha}\} + (n_{ij} - a'_{ij})\{\log \tfrac{1}{2} + \alpha(\log\theta_j - \log\theta_i)\} \right]$$
$$+ \sum_{i(<j)=1}^{t} \left[\log C_{a_{ij}}^{n_{ij}} + a_{ij} \log \tfrac{1}{2} + a_{ij}\alpha\left(\log\theta_i - \log\theta_j\right) + (n_{ij} - a_{ij}) \log\{1 - \tfrac{1}{2}(\theta_i/\theta_j)^{\alpha}\} \right],$$

or

$$l \propto \sum_{i(<j)=1}^{h} \left[a'_{ij} \log\{1 - \tfrac{1}{2}(\theta_j/\theta_i)^{\alpha}\} + (n_{ij} - a'_{ij})\{\alpha(\log\theta_j - \log\theta_i)\} \right]$$
$$+ \sum_{i(<j)=1}^{t} \left[a_{ij}\alpha\left(\log\theta_i - \log\theta_j\right) + (n_{ij} - a_{ij}) \log\{1 - \tfrac{1}{2}(\theta_i/\theta_j)^{\alpha}\} \right],$$

where $l = logL(\boldsymbol{a}; \boldsymbol{\theta})$ and $\boldsymbol{\theta} = (\theta_1, \theta_2, ..., \theta_t)$ be the set of *t* parameters designating the worth of the competing items under-study.

Taking the first partial derivatives of (4.1) with regard to θ_i and equating to zero, we get

$$\frac{\partial l}{\partial \theta_i} = \frac{\alpha}{\theta_i}\left[\sum_{i(<j)=1}^{h} a'_{ij} - \sum_{i(<j)=1}^{h} \frac{0.5a'_{ji}(\theta_j/\theta_i)^{\alpha}}{1-0.5(\theta_j/\theta_i)^{\alpha}} - \sum_{i(<j)=h+1}^{t} a_{ji} + \sum_{i(<j)=h+1}^{t} \frac{0.5a_{ji}(\theta_i/\theta_j)^{\alpha}}{1-0.5(\theta_i/\theta_j)^{\alpha}}\right] = 0,$$

for all $i = 1, 2, ..., t$, where $\boldsymbol{\theta}$, the vector of the unknown worth parameters θ_i, $i = 1, ..., t$, refers to the worth of item *i*, $C_{a_{ij}}^{n_{ij}}$ be the number of combinations of a_{ij} out of n_{ij} and the rest of the notations are self-defined.

4.1. Standard deviations of the ML estimates

Standard deviations of the ML estimates are on the main diagonal of the inverted information matrix. The information matrix is given by the expectation of the negative Hessian having (i, j) element $I(\boldsymbol{\theta})_{ij} = -E\left(\frac{\partial^2 \log L(\boldsymbol{\theta})}{\partial\theta_i\theta_j}\right)$. Since we just need to compute the variances, so we just need to evaluate $I(\boldsymbol{\theta})_{ii} = -E\left(\frac{\partial^2 l}{\partial\theta_i^2}\right)$ for all $i = 1,2, \dots, t$. Similarly, the other equation may have been derived but omitted for the sake of brevity. The off-diagonal elements $\frac{\partial^2 \log l}{\partial\theta_i\theta_j}$ are found to be:

$$\frac{\partial^2 \log l}{\partial\theta_i\theta_j} = \frac{2\alpha^2 a_{ji}\left(\theta_i/\theta_j\right)^{\alpha-1}}{\theta_j^2\left\{2-\left(\theta_i/\theta_j\right)^{\alpha}\right\}^2}, \quad \theta_i < \theta_j, \quad i(<j) = 1,2, \dots, t.$$

5. Numerical illustration

For illustration, we collected a real dataset on five top-ranked one-day-international cricket teams of Australia, India, New Zealand, Pakistan and South Africa from the website *www.howstat.com* and is given in Table 1. In the light of the dataset, the binomial variable a_{ij} attains the respective values 15, 12, 10, 15, 3, 9, 6, 6, 6, 3 for $i(<j) = 1, 2, ..., 5$; and 4, 6, 7, 4, 8, 11, 9, 7, 10, 6 for $i(>j) = 1, 2, ..., 5$ and zero for $i = j$.

Table 1: Observed Dataset of ODI Cricket Matches

Teams	Australia	India	New Zealand	Pakistan	South Africa
Australia	0	15	12	10	15
India	4	0	3	9	6
New Zealand	6	7	0	6	6
Pakistan	4	8	11	0	3
South Africa	9	7	10	6	0

To find solution to the likelihood equations $\frac{\partial logL\,(\boldsymbol{a};\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = 0$, we develop computer program in *SAS* software using *PROC SYSNLIN* procedure, the resulting ML estimates are displayed in Table 2.

Table 2: ML estimates for the observed dataset

Teams	Parameters	Estimates
Australia	θ_1	0.381424
India	θ_2	0.123290
New Zealand	θ_3	0.135773
Pakistan	θ_4	0.145787
South Africa	θ_5	0.213726

Here we see that the teams under-consideration may be ranked as follows: The Australians stand first, the South Africans are the second, Pakistanis being the third, New Zealanders being the fourth and finally Indians being the last and have lowest rank.

6. Prior distributions

Bayesian inference necessitates the use of certain prior distributions. Frequently used uninformative priors include the uninform and Jeffreys' priors. Informative distribution must represent the information about the parameters and, obviously, it must be in agreement with the parameter space.

6.1. The uninformative uniform prior

The uninformative uniform prior assigns equal weights to all of the items and may be defined as:

$$p(\boldsymbol{\theta}) = 1,\ \ 0 < \theta_i < \infty, i = 1, 2, ..., t.,$$

where $\boldsymbol{\theta} = (\theta_1, \theta_2, ..., \theta_t)$ is the vector of unknown parameters.

6.2. The uninformaitive Jeffreys' prior

The Jeffreys' prior for the vector of the unknown parameters $\boldsymbol{\theta}$ is $p(\boldsymbol{\theta}) \propto \sqrt{|I(\boldsymbol{\theta})|}$, where $I(\boldsymbol{\theta}) = -E\left(\frac{\partial^2 \log L(\boldsymbol{a};\boldsymbol{\theta})}{\partial \theta_i \theta_j}\right)$ is the Fisher's Information Matrix. E stands for the expectation on the dataset and i and j respectively denote the row and column of the determinant.

6.3. The informative Dirichlet prior

The nature and the range of parameters generally determine the prior distribution to be assigned to the parameters. For the present case, we have location parameter of the Pareto distribution as the parameter of interest with a support in the interval (0, ∞). But we have witnessed that maximum area of the parameter is condensed within the range (0, 1). Hence we have chosen a multivariate Dirichlet prior for all the parameters of interest θ_i, for $i = 1, 2, ..., t$, that is

$$p(\boldsymbol{\theta}) = \frac{\Gamma(a_1 + a_2 + ... + a_t)}{\Gamma(a_1)\Gamma(a_2)...\Gamma(a_t)} \prod_{i=1}^{t} \theta_i^{a_i - 1}, \quad 0 \le \theta_i \le 1, \quad \sum_{i=1}^{t} \theta_i = 1, \quad a_i > 0, \forall i = 1, 2, ..., t,$$

be the set of hyperparameters. After choosing the prior, one still needs to elicit the hyperparameters.

6.4. Elicitation of the Hyperparameters

Using the prior predictive distribution approach discussed in Abbas and Aslam (2009), we elicit the hyperparameters of the Dirichlet prior. Using the proposed model and the Dirichlet prior, the prior predictive distribution $p_{(ij)}$ for the variable a_{ij} is:

$$p_{(a_{ij})} = K_{ij} \int_0^1 \left[M \left\{\phi_{ij}\right\}^{a_{ij}} \left\{1 - \phi_{ij}\right\}^{a_{ji}} \right] d\theta_i, \quad \text{for } a_{ij} = 0, 1, ..., n_{ij},$$

where $M = (\theta_i)^{a_i - 1}(1 - \theta_i)^{a_j - 1}$, $\theta_j = 1 - \theta_i$, $n_{ij} = a_{ij} + a_{ji}$ and $K_{ij} = C_{a_{ij}}^{n_{ij}} / \{2^{a_{ij}} B(a_i, a_j)\}$.

Considering the real data set described in Section 5, we take at most 4 (say) preferences/wins of item i over j and use the predictive distribution $p_{(a_{ij})}$ to find the predicted probability of at most four wins as $p_{0(ij)} = P(a_{ij} \le 4) = \sum_{a_{ij}=0}^{4} p_{(a_{ij})}$. The corresponding elicited probability of the same number of preferences/wins (i.e., at most 4) of the item i over j is found using the binomial law as $P_{0(ij)} = P(a_{ij} \le 4) = \sum_{a_{ij}=0}^{4} B\left(a_{ij}; n_{ij}, \psi_{ij}\right)$, where n_{ij} (known) and ψ_{ij} (unknown). Here ψ_{ij} is the probability of success (preferring or winning) in single trial for all the team-pairs and is obtained from the cricket-experts. We may calculate ψ_{ij} based on the observed dataset given in Table 1 under the assumption that the preference probabilities furnished by cricket-experts will not be significantly different from those found using the observed dataset, that is, the preference probabilities ψ_{ij} are found using $\psi_{ij} \approx \phi_{ij} = a_{ij}/n_{ij}$ for $i(<j) = 1, 2, ..., 5$.

Now we need to find values of the hyperparameters $a_i, \forall i = 1, 2, ..., 5$, that minimize the difference between the fitted predictive probabilities $p_{(ij)}$ and the elicited probabilities $p_{0(ij)}$. To accomplish this, a program is written in *SAS* package using its *PROC SYSNLIN* procedure, and the elicited values of the hyperparameters are found to be a_1= 38.86711, a_2= 45.45095, a_3= 59.48605, a_4= 20.61834 and a_5= 92.21675 , and the desired elicited informative Dirichlet prior is:

$$p(\boldsymbol{\theta}) = \frac{\theta_1^{38.86711-1}\theta_2^{45.45095-1}\theta_3^{59.48605-1}\theta_4^{20.61834-1}\theta_5^{92.21675-1}}{\beta\left(38.86711, 45.45095, 59.48605, 20.61834, 92.21675\right)}, \quad 0 < \theta_i < 1,$$

for $i = 1, 2, .., 5$ and $\theta_5 = (1 - \theta_1 - \theta_2 - \theta_3 - \theta_4)$.

7. Posterior distribution

So far as the selection of prior is concerned, we may choose uninformative or informative priors. Using any of the above-stated priors and the likelihood function, the joint posterior distribution $p(\boldsymbol{\theta} \mid \mathbf{a})$ of the model parameters $\boldsymbol{\theta}$, conditional upon the observed dataset of Table 1, is

$$p(\boldsymbol{\theta} \mid \mathbf{a}) = K_U^{-1} \prod_{i(<j)=1}^{t} \left[p(\theta_i, \theta_j) \left\{\phi_{ij}\right\}^{a_{ij}} \left\{1 - \phi_{ij}\right\}^{a_{ji}} \right], \quad \theta_i > 0,$$

where $p(\theta_i, \theta_j)$ may be any of the chosen priors, and the normalizing constant is:

$$K_U = \int_{\boldsymbol{\theta}} \prod_{i(<j)=1}^{t} \left[p(\theta_i, \theta_j) \left\{\phi_{ij}\right\}^{a_{ij}} \left\{1 - \phi_{ij}\right\}^{a_{ji}} \right] d\boldsymbol{\theta}.$$

Notice that we impose the constraint $\sum_{i=1}^{t} \theta_i = 1$ for identification and use five items taking t =

5 throughout the analysis made in this study.

7.1. Marginal posterior distributions

To see the nature of variation of the individual worth parameters θ_i for $i = 1, 2, ..., 5$, we find their marginal posterior distributions based on the observed dataset. Such distribution for any parameter θ_i may be found by integrating out all the nuisance parameters. For instance, the marginal posterior distribution of the parameter θ_1 for the Australian team is

$$p(\theta_1|a) = \int_0^{1-\theta_1} \int_0^{1-\theta_1-\theta_2} \int_0^{1-\theta_1-\theta_2-\theta_3} p(\boldsymbol{\theta}|\boldsymbol{a}) d\theta_4 d\theta_3 d\theta_2 \, , \quad 0 \leq \theta_1 \leq 1.$$

The joint posterior distribution has a complicated expression and the marginal posterior distributions cannot be derived in closed form. However, we can make use of the numerical integration to evaluate it. Plotting the values of the worth parameters against their ordinates, we draw graphs to view the behavior of the worth parameters using the uniform prior. The desired graphs are given in Figure 2.

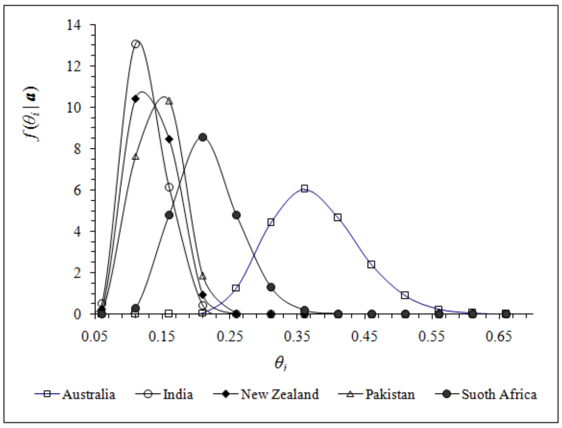

Figure 2: Marginal posterior distributions via the uniform prior

Here, we see that the marginal posterior distributions of the worth parameters have varying dispersions with highest ordinates at the values (joint posterior modes) close to the posterior means.

8. Bayesian analysis – An illustration

8.1. Posterior means

We first assume a squared error loss function and find the posterior means as the Bayes estimates of the worth parameter θ_i of the *i*th team as

$$E(\theta_i|\boldsymbol{a}) = \int_0^1 p(\theta_i|\boldsymbol{a})d\theta_i, \quad 0 \le \theta_i \le 1,$$

where $p(\theta_i\,|\,\mathbf{a})$, the marginal posterior distribution of the model parameter θ_i, is

$$p(\theta_i|\boldsymbol{a}) = \int_0^1 p(\boldsymbol{\theta}|\boldsymbol{a})d\boldsymbol{\theta}', \quad 0 \le \theta_i \le 1.$$

Here $\boldsymbol{\theta}'$ and θ_i form partition of $\boldsymbol{\theta}$ such that $\boldsymbol{\theta}' \cup \theta_i = \boldsymbol{\theta}$ and $\boldsymbol{\theta}' \cap \theta_i = \emptyset$. To evaluate the complicated integrals we use the Quadrature method of numerical integration, which refers to numerically approximating the value of a definite integral $\int_a^b p(\boldsymbol{\theta})d\boldsymbol{\theta}$. To accomplish this, we calculate it at a number of points in the range a to b and find the result as a weighted average as $\int_a^b p(\boldsymbol{\theta})d\boldsymbol{\theta} \cong \sum_{i=1}^{n} \varepsilon_i p(\boldsymbol{\theta})$, where ε_i denotes the increment used to b through a. Here it is important to note that accuracy of the integral is inversely proportional to the size of increment. A two-dimensional integration may be found as $\int\limits_{\theta_i}\int\limits_{\theta_j} p\left(\theta_i,\theta_j\right)d\theta_i d\theta_j \cong \sum_{i=o}^{n_i}\sum_{j=0}^{n_j}\varepsilon_i\varepsilon_j\, p\left(\theta_i,\theta_j\right)$, where the notations are pre-defined. Likewise the higher-order integrals may be evaluated. For our case, expression involves five parameters and integration seems intractable. So we take the scale parameter $\alpha = 0.55$ and assume the sum of parameters to be unity for identification. We also calculate the Jeffreys' prior by numerical differentiation and use it for the calculation of the posterior means of the parameters by numerical integration through the quadrature method by developing computer programs in in *C* and SAS languages and the resulting posterior means, their standard errors within parentheses, chi-square values and *p*-values are displayed in Table 3. The motivation for furnishing here the *p*-values and chi-square values is to test the hypothesis of the appropriateness of the proposed model under study. A detailed account on the appropriateness of the model is however given in Section 11.

Table 3: Bayes estimates via the uninformative and informative priors for the observed dataset

Teams	Parameters	Bayes Estimates		
		Uniform prior	Jeffreys' prior	Dirichlet prior
Australia	θ_1	0.37557 (0.06701)	0.36497 (0.065088)	0.21527 (0.01911)
India	θ_2	0.12614 (0.02900)	0.12070 (0.029535)	0.15812 (0.01910)
New Zealand	θ_3	0.13562 (0.03076)	0.13404 (0.032242)	0.19190 (0.01884)
Pakistan	θ_4	0.14579 (0.03228)	0.14279 (0.033347)	0.11621 (0.01440)
South Africa	θ_5	0.21688 (0.04695)	0.23751 (0.053079)	0.31850 (0.02427)
Chi-square values		3.96834	4.17025	20.45579
p-values		0.68096	0.65365	0.00230

From the results we observe that for the uninformative priors, the Australia stands first, South Africa the second, Pakistan the third, New Zealand the fourth and India the fifth and last one. We observe that the posterior estimates under the uniform and the Jeffreys' priors substantially agree. However, the Bayes estimates using the informative prior reflect a change in ranking order of the teams under-study. South Africa is now on the top, Australia the second, New Zealand the third, India the fourth and Pakistan being the fifth and the last one. It reveals that a potential prior information may change the inference about the parametric values. It may further be added that the observed insignificant *p*-values and chi-square values for all the uninformative priors establish the appropriateness of the proposed model.

8.2. Preference probabilities

Probabilities indicating the expected probabilities of preferences of any item over the other in any future contest or comparison are termed as the preference probabilities. Such probabilities are calculated using the proposed model for all the priors and its estimates are displayed in Table 4. Here the team names are coded for brevity as 'AU' for Australia, 'IN' for India, 'NL' for New Zealand, 'PA' for Pakistan and 'SA' for South Africa.

Table 4: Preference probabilities ϕ_{ij} for the observed dataset

Team-Pairs	Uniform prior	Jeffreys' prior	Dirichlet prior
(AU, IN)	0.72562	0.72794	0.57804
(AU, NL)	0.71446	0.71179	0.53062
(AU, PA)	0.70287	0.70159	0.64378
(AU, SA)	0.63033	0.60522	0.40309
(IN, NL)	0.48046	0.47199	0.44949
(IN, PA)	0.46173	0.45585	0.57790
(IN, SA)	0.37112	0.34458	0.34018
(NL, PA)	0.48050	0.48291	0.62054
(NL, SA)	0.38621	0.36503	0.37840
(PA, SA)	0.40188	0.37795	0.28717

Obviously the preference probabilities are compatible with the posterior means of the competing teams regarding the ranking order established therein.

8.3. Bayesian hypotheses testing

We define null and the alternative hypotheses as

$$H_{ij}:\theta_i > \theta_j \text{ vs } H_{ji}:\theta_i \le \theta_j, \forall\, i(<j) = 1, 2, 3, 4, 5.$$

The posterior probabilities p_{ij} and q_{ij} of the respective hypotheses H_{ij} and H_{ji} are calculated using the density $p\left((\phi_{12},\xi_1,\theta_3,\theta_4,\theta_5)\,|\,\mathbf{a}\right)$ derived by reparameterization in the joint posterior $p(\boldsymbol{\theta}\,|\,\mathbf{a})$ as $\phi_{ij} = \theta_i - \theta_j$ and $\xi_i = \theta_i$, $\forall i(<j) = 1, 2, 3, 4, 5$. The following rule applies to draw conclusion about the hypotheses regarding the teams being compared. Let

$$s = \min(p_{ij}, q_{ij}),$$

if p_{ij} (or q_{ij}) is large, then H_{ij} (or H_{ji}) is accepted with high probability. This implies that if 's' is small, we can reject one of the hypotheses, otherwise if $s > 0.1$ (say), then the evidence is inconclusive. Being specific, the posterior probability p_{12} of the hypothesis $H_{12}:\theta_1 > \theta_2$ that Australians will defeat India, is evaluated as

$$p_{12} = p\left(\phi_{12} > 0\,|\,\mathbf{a}\right) = \int_0^1\int_\phi^{(1+\phi)/2}\int_0^{1-2\xi+\phi}\int_0^{1-2\xi+\phi-\theta_3} p\left((\phi_{12},\xi_1,\theta_3,\theta_4,\theta_5)\,|\,a_{ij}\right) d\theta_4 d\theta_3 d\xi_1 d\phi_{12}.$$

Here, as usual, the subscripts 1 through 5 represent the Australians, Indians, New Zealanders, Pakistanis and South Africans respectively. The posterior probability of the alternative hypothesis $H_{21}:\theta_1 \le \theta_2$ is $q_{12} = 1 - p_{12}$.

The posterior probabilities are evaluated to be as hight as very close to one in favor of the hypotheses consistent with the ranking order established via the ML and the Bayes estimates.

8.4. Predictive probabilities

Having observed sample $\boldsymbol{a}$ for the number of preferences of item *i* over *j*, Bayesians use the posterior predictive distribution $p(a'_{ij} \mid \mathbf{a})$ to predict the future observation a'_{ij} as:

$$p(a'_{ij}|\boldsymbol{a}) = \int_{\boldsymbol{\theta}} P(a'_{ij}|\boldsymbol{\theta})p(\boldsymbol{\theta}|\boldsymbol{a})d\boldsymbol{\theta}, \quad a'_{ij} = 0,1,\ldots,n_{ij}$$

where $p(a'_{ij} \mid \boldsymbol{\theta})$ is the property of the model, and distributions of the variables $\boldsymbol{a}$ and a'_{ij} are fully characterized by $\boldsymbol{\theta}$. The predictive probability $P_{(ij)}$ that item *i* will be preferred to *j* in a future single comparison is

$$P_{(ij)} = \int_{\boldsymbol{\theta}} p(T_i > T_j \mid \boldsymbol{\theta})\, p(\boldsymbol{\theta} \mid \mathbf{a})d\boldsymbol{\theta},$$

where $p(T_i > T_j \mid \boldsymbol{\theta})$ is given by a PC model. The predictive probability of winning of any team against another in a single future contest is denoted by $P_{(ij)}$, for $i(< j) = 1,\ldots,5$, and is found as

$$P_{(ij)} = \int_{\boldsymbol{\theta}} \phi_{ij}\, p(\boldsymbol{\theta} \mid \mathbf{a})d\boldsymbol{\theta}, \quad i(< j) = 1, 2, 3, 4, 5.$$

where ϕ_{ij} stand for the winning (preference) probability of the team *i* against *j* and $p(\boldsymbol{\theta} \mid \mathbf{a})$ denotes the posterior distribution $\boldsymbol{\theta}$ conditional upon dataset $\mathbf{a}$. The posterior predictive probabilities $P_{(ij)}$ are evaluated via the numerical integration, and are furnished in Table 5.

Table 5: Pair-wise posterior $p_{(ij)}$ predictive probabilities for the observed dataset

Team-Pairs	Uniform prior	Jeffreys' prior	Dirichlet prior
(AU, IN)	0.72210	0.72339	0.57709
(AU, NL)	0.71068	0.70690	0.52925
(AU, PA)	0.71068	0.69634	0.64297
(AU, SA)	0.62557	0.59845	0.37786
(IN, NL)	0.48700	0.47919	0.45090
(IN, PA)	0.45188	0.44451	0.57577
(IN, SA)	0.37721	0.35101	0.34074
(NL, PA)	0.48606	0.48881	0.66211
(NL, SA)	0.39225	0.37149	0.37925
(PA, SA)	0.40867	0.38516	0.28764

Obviously the predictive probabilities support the ranking order of the teams under study already established via the posterior means and preference probabilities. However, the changes in the results found using the informative Dirichlet prior are due to the utility of additional information furnished by the cricket-experts.

9. Simulation study based on different configurations of parametric values

To assess the sustainability and performance of the proposed model, a simulation study is conducted considering a balanced paired comparison experiment each consisting of 50 trials and taking 100,000 simulations and using different configurations of parametric values. To reflect different populations, we have used equal and unequal (increasing and decreasing) parametric values to simulate 100,000 datasets from these populations. The posterior means of the associated worth parameters and their standard deviations are computed based on the simulated datasets via the uniform prior and are displayed in Table 6.

Table 6: Bayes estimates for datasets simulated for different parametric values using uniform prior

Teams	Parameters	Results of 100,000 simulations					
		Assumed equal values	Estimated values	Assumed increasing values	Estimated values	Assumed decreasing values	Estimated values
Australia	θ_1	0.20000	0.19566 (0.01829)	0.0835	0.15726 (0.01123)	0.3147	0.24962 (0.02685)
India	θ_2	0.20000	0.19125 (0.01854)	0.1324	0.17232 (0.01597)	0.2235	0.20417 (0.02128)
New Zealand	θ_3	0.20000	0.19685 (0.01895)	0.2054	0.20240 (0.01924)	0.1375	0.17421 (0.01729)
Pakistan	θ_4	0.20000	0.20192 (0.01903)	0.3223	0.23416 (0.01911)	0.0579	0.12743 (0.01352)
South Africa	θ_5	0.20000	0.21432 (0.02052)	0.2564	0.23386 (0.02079)	0.2664	0.24457 (0.02575)

From the results reflected in Table 6, it becomes evident that the Bayes estimates and the corresponding true parametric values coincide a lot. So it is established that the estimation technique utilized therein is efficient. Moreover, the preference probabilities ϕ_{ij} are also evaluated for all the team pairs on behalf of the posterior parametric estimates obtained using the simulated datasets in the three assumed-value configurations for the proposed model and are presented in Table 7.

Table 7: Preference probabilities ϕ_{ij} based on the estimates for the simulated datasets

Team-Pairs	For estimates based on assumed equal values	For estimates based on assumed increasing values	For estimates based on assumed decreasing values
(AU, IN)	0.50623	0.47421	0.55233
(AU, NL)	0.49833	0.42556	0.58974
(AU, PA)	0.49126	0.37761	0.65457
(AU, SA)	0.47431	0.37805	0.50559
(IN, NL)	0.49213	0.45766	0.54560
(IN, PA)	0.48529	0.42240	0.64799
(IN, SA)	0.46964	0.42270	0.45273
(NL, PA)	0.49306	0.46148	0.59383
(NL, SA)	0.47716	0.46181	0.41490
(PA, SA)	0.48388	0.50035	0.34934

Obviously, the preference probabilities second the ranking order established via the posterior parametric estimates obtained using the simulated datasets.

10. Comparison of the priors

The uninformative uniform and Jeffreys' priors may be compared on the basis of the estimates produced by them. Here we see that the posterior means found using the uniform and Jeffreys' priors agree up to one decimal place. However the uninformative Jeffreys' and the informative Dirichlet priors may be compared on the basis of the Lindly-Shannon information (LSI). In accordance with the definition of information proposed by Lindley (1956), if $p(\boldsymbol{\theta})$ denotes prior distribution about a state of nature $\boldsymbol{\theta}$, then the amount of information, $I\{p(\boldsymbol{\theta}\}$, contained in this prior is defined as

$$I\{p_J(\boldsymbol{\theta})\} = \int_{\boldsymbol{\theta}} p_J(\boldsymbol{\theta}) \ln\{p_J(\boldsymbol{\theta})\} d\boldsymbol{\theta},$$

provided that the integral converges. The amounts of the respective LSIs for the uninformative Jeffreys' and the informative Dirichlet priors are found to be 5.7224×10^{53} and 4.6216×10^{160}. Here it may be said in the present environment that the Jeffreys' prior is less informative than the Dirichlet prior.

11. Appropriateness of the model

A model is said to give appropriate fit to the observed dataset if the expected frequencies obtained using the model are substantiallly close to the observed frequencies. To test the hypothesis of the appropriateness of a model, we evaluate χ^2-values. The smaller the value of χ^2, the better the fit would be. Let the ordered pairs $(a_{ij}, \hat{a}_{ij})$ and $(a_{ji}, \hat{a}_{ji})$ respectively denote the observed and the corresponding expected frequencies for the preferences of item *i* over *j* and the vice versa. Then the χ^2-statistic is defined as

$$\chi^2 = \sum_{i<j}^{t} \left\{ \frac{(a_{ij} - \hat{a}_{ij})^2}{\hat{a}_{ij}} + \frac{(a_{ji} - \hat{a}_{ji})^2}{\hat{a}_{ji}} \right\},$$

with $(t-1)(t-2)/2$ degrees of freedom [Stern (1990), Aslam, (1995), Abbas and Aslam (2010b)]. The null hypothesis H_0 and the alternative hypothesis H_1 about the values of the model parameters are

H_0 : The model is true for the values of parameter $\boldsymbol{\theta}$.

H_1: The model is not true for any values of the parameters.

The expected frequencies are calculated as $\hat{a}_{ij} = n_{ij}.\phi_{ij}$ and $\hat{a}_{ji} = n_{ij}.(1-\phi_{ij})$. The statistical measures, that is, the χ^2 and the associated *p*-values are obtained for different values of the α and are presented in Table 8.

Table 8: χ^2-values with associated *p*-values for different α-values

α-values	Goodness indicators	Classical ML estimates	Bayes estimates via the uniform prior
0.05	χ^2-values	98.54342	3.97641
	p-values	0.00000	0.67987
0.55	χ^2-values	3.94733	3.96834
	p-values	0.68380	0.68096
1.00	χ^2-values	3.94732	4.01918
	p-values	0.68380	0.67408
2.00	χ^2-values	3.94732	4.06574
	p-values	0.68381	0.66778
5.00	χ^2-values	3.94731	4.16597
	p-values	0.68381	0.65423
10.00	χ^2-values	3.94731	4.91929
	p-values	0.68381	0.55421
50	χ^2-values	3.94740	17.64618
	p-values	0.68380	0.00718
100	χ^2-values	3.94747	17.64618
	p-values	0.68379	0.00718

From the results it reveals that there is no evidence to conclude that the proposed Pareto model under consideration does not fit to the observed dataset. Moreover, the *p*-values associated with the maximum likelihood estimates and those found using the uniform prior are observed to be the highest ones, so these may be declared the most accurate estimates. Moreover, highest *p*-values are observed to exist for $\alpha = 0.55$, hence this value is proved to be the best choice.

12. Concluding remarks with discussion

Paired comparison models are proposed using the Pareto distribution. Whereas, we have estimated the best value of the scale parameter via the sensitivity analysis. Having a view of the facts and figures of the analysis given in the form of the ML and Bayes estimates, we notice that the five cricket teams under study may be ranked as Australia being the number one, South Africa the second one, Pakistan being the third one, New Zealand the fourth and India being the fifth and last one for the uninformative priors. It is also worth mentioning that the estimates obtained in the forms of the posterior means under the Uniform and the Jeffreys' priors agree considerably. The posterior estimates via the informative prior reflect a different ranking order due to incorporating some additional information. The preference probabilities, posterior predictive probabilities and the posterior probabilities of hypotheses also favor the same ranking order.

It has also been observed that the Bayes estimates based on the simulated datasets confirm the ranking order of the teams under study that was observed using the real dataset and generate very small standard errors indicating the best performance of the model in the theoretical conditions. It is also observed that the Bayes estimates and the corresponding true parametric values coincide a lot for datasets simulated in different populations. So it is established that the estimation technique utilized therein is efficient. The test for goodness of fit of the model conducted through the χ^2-statistic also indicates the appropriateness of model. We have witnessed a variation in the inferences drawn using the uninformative and the informative priors, which indicates that a potential prior information may affect the inferences based on the priors.

References

Abbas, N. and Aslam, M. (2010). Bayesian analysis of gamma models for paired comparisons. *Pakistan Journal of Statistics*, **26**(3), 479-491.

Abbas, N. and Aslam, M. (2011a). Extending the Bradley-Terry Model for Paired

Comparisons to Accommodate Weights. *Journal of Applied Statistics*, **38**(3), 571-580.

Abbas, N., Aslam, M. and Hussain, Z. (2011b). Refining the Judges' Assessments of Items in Paired Comparison Experiments. *Hacettepe Journal of Mathematics and Statistics*, **40** (6), 907 – 913.

Abbas and Aslam (2012a). Modeling the Factors influencing the Evaluations of Items through Mixture Models for Preference Datasets. *Communications in Statistics - Theory and Methods* **42**(1), 135-144.

Abbas, N. and Aslam, M. (2013). Eliciting Hyperparameters of Prior Distributions for the Parameters of the Paired Comparison Models. *Pakistan Journal of Statistics and operations research,* **9**(1), 71-78.

Aslam, M. (1995). Bayesian analysis for paired comparison data. Ph. D. Thesis, University of Wales, Aberystwyth, U.K.

Bradley, R.A. and M.E. Terry (1952). Rank Analysis of Incomplete Block Designs: I. The Method of Paired Comparisons, *Biometrika*, 39, 324–345.

David, H. A. (1988). "The Method of paired comparisons," second ed. Charles Griffin and Company Ltd. London.

Davidson, R.R. (1970). On extending the Bradley-Terry model to accommodate ties in paired comparison experiments. *Journal of the American Statistical Association* **65**, 317–328.

Degroot, M. H. (1970). Optimal Statistical Decisions. McGraw-Hill Book Company, New York.

Glenn, W.A. and David, H. A. 1960. Ties in Paired Comparison experiments using a modified Thurstone-Mosteller Model. *Biometrics* **16**, 86–109.

Henery, R. J. (1981). Permutation probabilities as models for horse races. *Journal of the Royal Statistical Society* B **43**, 86–91.

Henery, R. J. (1986). Interpretation of average ranks. *Biometrika* **73**, 224–227.

Lindley, D. V. (1956). On a measure of information provided by an experiment.*Annals Of Mathematical Statistics*, **27**, 986-1005.

Logan, R.J. (2012). Forecasting, When power law distributions apply. Master thesis, University in Wellington, New Zealand.

Mak, T. K. (1985). Interpretation and estimation in ranking theory. *Biometrika* **77**, 227–229.

Rao, P.V. and Kupper, L.L.(1967). Ties in paired-comparison experiments: a generalization of the Bradley-Terry model. *Journal of the American Statistical Association* **62**, 194–204.

Stern, H. (1990). A continuum of paired comparison models. Biometrika 77, pp. 265-273.

Thompson, W. A. AND Singh, J. (1967). The use of limit theorems in paired comparison model building. *Psychometrica* **32**, 255–264.

Thurstone, L. L. (1927). A law of comparative judgment. *Psychological Review* **34**, 273–286.