▼ Journal of Research (Science), Bahauddin Zakariya University, Multan, Pakistan.
Vol.14, No.2, December 2003, pp. 159-167 ISSN 1021-1012

ON THE POSTERIOR DISTRIBUTION OF PARAMETERS FOR THE RAO-KUPPER MODEL

M. Aslam
Department of Statistics, Quaid-i-Azam University, Islamabad, 44000, Pakistan.
email: aslamsdqu@yahoo.com

Abstract: Posterior distribution is the workbench in Bayesian analysis. The posterior distribution for the parameters of the Rao-Kupper model for paired comparison data is derived using the (informative) Dirichlet-gamma prior distribution. In this study, an attempt is made to find analytical expression for the posterior (marginal) distributions of the parameters of the Rao-Kupper model. Five approximate analytical expressions for the marginal posterior distributions of the threshold parameter and three analytical expressions for posterior distribution of the treatment parameters are presented in this study.

Keywords: Dirichlet-gamma distribution, paired comparison method, prior and posterior distributions, Rao-Kupper model.

INTRODUCTION

Sometime 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 sometime regarded as more reliable and can be obtained more readily from panelists. A detailed review of the literature concerning the methods/models of paired comparisons has been given by David [1988].

Consider a paired comparison trial with 'm' treatments or objects $T_1, T_2, ..., T_m$. There are 'm(m-1)/2' all possible pairs of 'm' treatments. Each pair (T_i, T_j), $(i < j, 1 \le i, j \le m)$ is ranked r_{ij} times independently. Let $\theta_1, \theta_2, ..., \theta_m$ 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 (a latent variable is a quantity which either in practice or in principle, cannot be directly observed or measured). The assumed mechanism is that she prefers treatment T_i to treatment T_j if $X_i > X_j$. Bradley and Terry [1952], and Bradley [1953] proposed a model of paired comparisons for the trial mentioned above. The Bradley-Terry model implies that the difference between two latent variables $(X_i - X_j)$ has a logistic (squared hyperbolic secant) density with location parameter $(\ln\theta_i - \ln\theta_j)$. Let ψ_{ij} denotes the probability

$\mathbf{P}(X_i > X_j | \theta_i, \theta_j)$ that the treatment T_i is preferred to the treatment T_j $(i \neq j)$ when the treatments T_i and T_j are compared then

$$\psi_{ij} = \frac{1}{4}\int_{-(\ln\theta_i - \ln\theta_j)}^{\infty} \operatorname{sech}^2(y/2)dy$$

$$= \frac{\theta_i}{\theta_i + \theta_j}, \quad i \neq j,\ i,j=1,2,...,m. \tag{1}$$

The Bradley-Terry model defined by (1) does not allow ties.

THE RAO-KUPPER MODEL FOR PAIRED COMPARISONS

Rao and Kupper [1967] proposed a modification of the Bradley-Terry model to allow for tied observations. They introduced a threshold parameter $\delta = \ln\lambda$ and assumed that if the observed difference $(X_i - X_j)$ is less than δ then the panelist is unable to distinguish between T_i and T_j and will declare a tie. Now the probability $\mathbf{P}\{(X_i - X_j) > \delta | \theta_i, \theta_j\}$ 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 denoted by $\psi_{i.ij}$ and is given by:

$$\psi_{i.ij} = \frac{1}{4}\int_{-(\ln\theta_i - \ln\theta_j)+\delta}^{\infty} \operatorname{sech}^2(y/2)dy$$

$$= \frac{\theta_i}{\theta_i + \lambda\theta_j} \tag{2}$$

The probability that treatment T_j is preferred to treatment T_i is denoted by $\psi_{j.ij}$ and may be obtained by swapping i with j in (2). The probability that treatments T_i and T_j have no preference is denoted by $\psi_{o.ij}$ and is given by

$$\psi_{o.ij} = \frac{1}{4}\int_{-(\ln\theta_i - \ln\theta_j)-\delta}^{-(\ln\theta_i - \ln\theta_j)+\delta} \operatorname{sech}^2(y/2)dy$$

$$= \frac{(\lambda^2 - 1)\theta_i\theta_j}{(\theta_i + \lambda\theta_j)(\lambda\theta_i + \theta_j)} \tag{3}$$

The Rao-Kupper model is given by Eqs. (2) and (3). If $\lambda=1$ then the Rao-Kupper model yields the Bradley-Terry model.

The notations to describe the data and the likelihood function are:

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

$n_{o.ijk}$ = 1 or 0 according as treatment T_i is tied with treatment T_j or not.

It is noted that: $n_{o.ijk} + n_{i.ijk} + n_{j.ijk} = 1$ and $n_{i.ijk} = n_{i.jik}$.

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

$n_{o.ij} = \sum_k n_{o.ijk}$ = number of times treatments T_i and T_j are tied.

r_{ij} **=** number of times treatment T_i is compared with treatment T_j and

$$r_{ij} = n_{o.ij} + n_{i.ij} + n_{j.ij} = r_{ji}.$$

The following notation is useful for further simplification of the likelihood,

$$n_{ijk} = n_{o.ijk} + n_{i.ijk}, \qquad n_{jik} = n_{o.ijk} + n_{j.ijk} = r_{ij} - n_{i.ijk},$$

$n_{ij} = \sum_k n_{ijk}$ = number of times treatment T_i is preferred to treatment T_j and number of times treatments T_i and T_j are tied,

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

$n_o = \sum_{i<j}^{m} n_{o.ij}$ = total number of times treatments T_i and T_j are tied.

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

$$P_{ijk} = (\lambda^2 - 1)^{n_{o.ijk}} \left(\frac{\theta_i}{\theta_i + \lambda\theta_j} \right)^{n_{ijk}} \left(\frac{\theta_j}{\lambda\theta_i + \theta_j} \right)^{n_{jik}}.$$

Now we put a constraint on the treatment parameters of the model that they are positive and sum to unity, this condition ensures that the parameters are well defined i.e. identifiable. Hence, the likelihood function of the observed outcome **x** {where **x** represents the data $(n_{i.ij}, n_{j.ij}, n_{o.ij})$} of the trial is:

$$l(\mathbf{x};\lambda,\theta_1,\ldots,\theta_m) = \prod_{i<j=1}^{m} \prod_{k=1}^{r_{ij}} P_{ijk}$$

$$= \frac{(\lambda^2 - 1)^{n_o} \prod_{i(<j)=1}^{m} K_{ij} \prod_{i=1}^{m} \theta_i^{n_i}}{\prod_{i \neq j}^{m} (\theta_i + \lambda\theta_j)^{n_{ij}}}, \tag{4}$$

where $K_{ij} = r_{ij} / (n_{o.ij}! n_{i.ij}! n_{j.ij}!)$, $0 \leq \theta_i \leq 1, (i = 1,2,\ldots,m), \sum_{i=1}^{m} \theta_i = 1$ and $\lambda > 1$, here λ is the threshold parameter and $\theta_1,\ldots,\theta_m$ are the treatment parameters.

THE CHOICE OF PRIOR DENSITY

One of the main differences between classical statistics and Bayesian statistics is that the latter can utilize the prior information in a formal way. The Dirichlet distribution is assumed to be a prior distribution for the treatment parameters $\theta_1, \theta_2, ..., \theta_m$ and for the threshold parameter λ the prior distribution is assumed to be independent of that for treatment parameters $\theta_1, \theta_2, ..., \theta_m$ and to have a gamma distribution, so the prior distribution of the parameters $\theta_1, \theta_2, ..., \theta_m$ and λ has the following form:

$$p(\lambda, \theta_1, ..., \theta_m) = \frac{e^{-b_2(\lambda-1)}(\lambda-1)^{b_1-1} b_2^{b_1}}{B(a_1, ..., a_m)\Gamma(b_1)} \prod_{i=1}^{m} \theta_i^{a_i-1}, \qquad (5)$$

$$0 \le \theta_i \le 1, (i = 1,2,...,m), \sum_{i=1}^{m} \theta_i = 1, \lambda > 1.$$

where $a_i (i = 1,...,m)$, b_1, b_2 are the hyperparameters and $a_i, b_1, b_2 > 0$.

THE POSTERIOR DISTRIBUTION FOR THE PARAMETERS OF THE MODEL

The posterior distribution for the treatment parameters $\theta_1, \theta_2, ..., \theta_m$ and threshold parameter λ using the likelihood function (4) and the prior distribution (5) is:

$$p(\theta_1, ..., \theta_m, \lambda | \mathbf{x}) \propto \frac{e^{-b_2(\lambda-1)}(\lambda-1)^{b_1+n_o-1}(\lambda+1)^{n_o} \prod_{i=1}^{m} \theta_i^{a_i+n_i-1}}{\prod_{i\neq j}^{m} (\theta_i + \lambda\theta_j)^{n_{ij}}}, \qquad (6)$$

$$\sum_{i=1}^{m} \theta_i = 1, 0 \le \theta_i \le 1, \lambda > 1,$$

where n_o, n_i and n_{ij} represent data.

AN ATTEMPT FOR CLOSED FORMS OF THE MARGINAL POSTERIOR DISTRIBUTIONS

The (marginal) posterior densities of the threshold parameter λ and the treatment parameter θ_i are derived and examined to find closed forms.

(a) The marginal posterior density of the threshold parameter λ for the case of two treatments has the following form by using the constraint $\theta_1 + \theta_2 = 1$, and taking $\theta = \theta_1 \Rightarrow \theta_2 = 1 - \theta$:

$$p(\lambda | \mathbf{x}) \propto e^{-b_2(\lambda-1)}(\lambda-1)^{b_1+n_{o.12}-1}(\lambda+1)^{n_{o.12}} \int_0^1 \frac{\theta^{a_1+n_{12}-1}(1-\theta)^{a_2+n_{21}-1} d\theta}{\{\theta + \lambda(1-\theta)\}^{n_{12}} \{\lambda\theta + (1-\theta)\}^{n_{21}}}$$

$$\lambda > 1, \qquad (7)$$

The integral in (7) is intractable; the following attempt is made to obtain the approximate analytical solution (closed form) for the integral of (7).

APPROXIMATION (I)

Let the integrand in (7) be denoted by $f(\theta)$ i.e.

$$f(\theta)=\frac{\theta^{a_1+n_{12}-1}(1-\theta)^{a_2+n_{21}-1}}{\{\theta+\lambda(1-\theta)\}^{n_{12}}\{\lambda\theta+(1-\theta)\}^{n_{21}}}, \qquad 0\le\theta\le 1. \qquad (8)$$

To draw different graphs, the hypothetical data given in following Table is considered and the values of the hyperparameters are assumed to be: $a_1=3$, $a_2=3$, $b_1=2$ and $b_2=1$.

Table: Hypothetical Data

Pair	$n_{1.12}$	$n_{2.12}$	$n_{o.12}$	n_{12}	n_{21}
1,2	10	30	01	11	31

The quadrature method is applied for the evaluation of the integrals. The shape of $f(\theta)$ has been examined at different values of λ (graph for λ=2 is given in Fig. 1). The graph looks like the beta distribution.

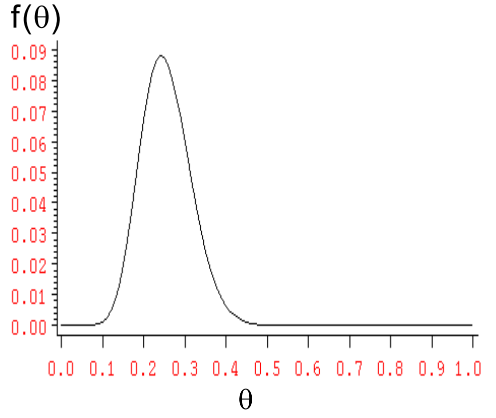


Fig. 1: Integrand of Posterior (Marginal) Density i.e. $f(\theta)$

In the denominator of $f(\theta)$, θ is replaced by (1-θ) in the first part of the product and (1-θ) by θ in second part, then the density of λ becomes:

$$p(\lambda|\mathbf{x})\propto e^{-b_2(\lambda-1)}(\lambda-1)^{b_1+n_{o.12}-1}(\lambda+1)^{n_{o.12}}\int_0^1\frac{\theta^{a_1+n_{12}-1}(1-\theta)^{a_2+n_{21}-1}d\theta}{(1+\lambda)^{n_{12}}(\lambda+1)^{n_{21}}\theta^{n_{21}}(1-\theta)^{n_{12}}}$$

$$\propto e^{-b_2(\lambda-1)}(\lambda-1)^{b_1+n_{o.12}-1}(\lambda+1)^{-r_{12}}B(a_1+n_{12}-n_{21},\ a_2+n_{21}-n_{12}),\ \lambda>1. \quad (9)$$

It is found that for large value of either n_{12} or n_{21} the density (9) becomes improper. So this approximation is not useful.

APPROXIMATION (II)

Now we try to replace the denominator of $f(\theta)$ by

$$h(\theta|\alpha_{(\lambda)},\beta_{(\lambda)}) = \theta^{\alpha_{(\lambda)}}(1-\theta)^{\beta_{(\lambda)}}.$$

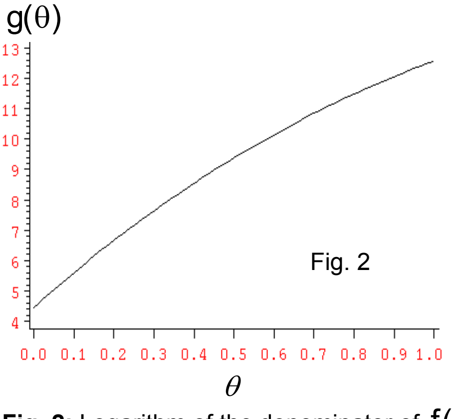
Then the density (7) is:

$$p(\lambda|\mathbf{x}) \propto e^{-b_2(\lambda-1)}(\lambda-1)^{b_1+n_{o.12}-1}(\lambda+1)^{n_{o.12}} \int_0^1 \frac{\theta^{a_1+n_{12}-1}(1-\theta)^{a_2+n_{21}-1}d\theta}{\theta^{\alpha_{(\lambda)}}(1-\theta)^{\beta_{(\lambda)}}}$$

$$\propto e^{-b_2(\lambda-1)}(\lambda-1)^{b_1+n_{o.12}-1}(\lambda+1)^{n_{o.12}} B\left(a_1+n_{12}-\alpha_{(\lambda)}, a_2+n_{21}-\beta_{(\lambda)}\right), \ \lambda>1. \quad (10)$$

where B stands for a Beta function and values of $\alpha_{(\lambda)}$ and $\beta_{(\lambda)}$ at various values of λ and θ (e.g. λ=1.5, θ =0.2 and 0.8) are calculated by the following equation:

$$\alpha_{(\lambda)}\ln(\theta)+\beta_{(\lambda)}\ln(1-\theta) \cong n_{12}\ln\{\theta+\lambda(1-\theta)\}+n_{21}\ln\{\lambda\theta+(1-\theta)\} \quad (11)$$



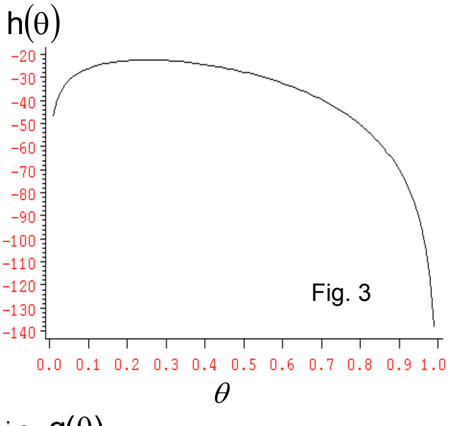
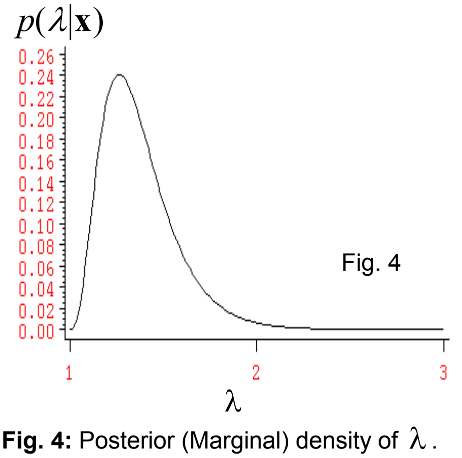


Fig. 2: Logarithm of the denominator of $f(\theta)$ i.e. $g(\theta)$.

Fig. 3: Transformed denominator i.e. $h(\theta)$.



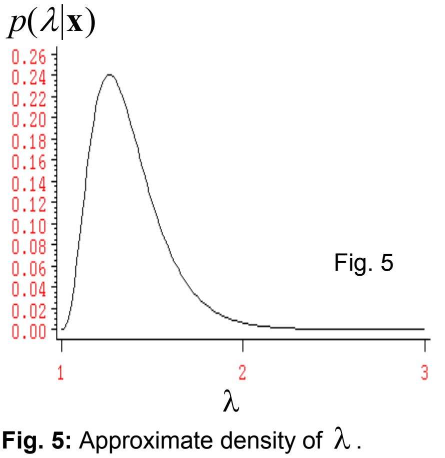


Fig. 4: Posterior (Marginal) density of λ. **Fig. 5:** Approximate density of λ.

The graphs of the right hand side {i.e. log of the denominator of $f(\theta)$} and left hand side {i.e. log of $h(\theta|\alpha_{(\lambda)},\beta_{(\lambda)})$} of (11) are given in Figs. 2 and 3 respectively. In this approximation, the density (10) remains proper. The graphs of (7) and (9) are given in Figs. 4 and 5 respectively for comparison. It is noted that they are approximately equal. This approximation is encouraging and can be used for analysis.

APPROXIMATION (III)

Another approximation is considered by examining the graph for the log of denominator of $f(\theta)$ (Fig. 2). The graph indicates an exponential function. So the denominator of the $f(\theta)$ is replaced by the exponential function $e^{-u\theta}$ then the integrand becomes:

$$f(\theta)=e^{-u\theta}\theta^{a_1+n_1-1}(1-\theta)^{a_2+n_2-1}, \quad \text{u is constant.} \tag{12}$$

Now the integral has a following closed form [Erdelyi *et al.* 1954]:

$$\int_0^1 e^{-\mu\theta}\theta^{c_1-1}(1-\theta)^{c_2-1}d\theta = B(c_1,c_2)\Phi_1(c_1,0,c_1+c_2,0,-\mu). \tag{13}$$

where $c_1 = a_1 + n_{12}, c_2 = a_2 + n_{21}$.

This approximation is not simple because the function Φ_1 is complicated.

APPROXIMATION (IV)

Let us consider λ = 1+ϕ where ϕ is small with high probability. Now the integral has a closed form as a Beta function but is suitable only for the small values of ϕ i.e. ϕ<0.1. The closed form of the integral of (7) is:

$$B(x,y)-n_{21}\phi B(x+1,y)-n_{12}\phi B(x,y+1)+.... \tag{14}$$

where $x = a_1 + n_{12}$, $y = a_2 + n_{21}$ and B stands for a Beta function as:

$$B(x,y)=\int_0^\infty \frac{z^{x-1}dz}{(1+z)^{x+y}}, \quad \text{where } z=\frac{\theta}{1-\theta}. \tag{15}$$

APPROXIMATION (V)

Other approximations like Lindley [1980] and Tierney-Kadane [1986] approximation can be employed to obtain the marginal posterior density of λ numerically because the modes of the parameters involved in both the approximations are not available in closed forms. Therefore, these approximations need to be done numerically.

(b) As $\theta_m = 1-\sum_{i=1}^{m-1}\theta_i$ so there are '*m*-1' effective treatment parameters, now the (marginal) posterior density of the treatment parameter θ_1 has the following form,

$$p(\theta_1|\mathbf{x}) \propto \int_{\theta_{m-1}=0}^{1-\theta_1} \cdots \int_{\theta_2=0}^{1-\sum_{i\neq 2}^{m-1}\theta_i} \int_{\lambda=1}^{\infty} \frac{e^{-b_2(\lambda-1)}(\lambda-1)^{b_1+n_o-1}(\lambda+1)^{n_o}\prod_{i=1}^{m}\theta_i^{a_i+n_i-1}}{\prod_{i\neq j}^{m}\left(\theta_i+\lambda\theta_j\right)^{n_{ij}}}$$

$$d\lambda d\theta_2,\ldots,d\theta_{m-1},\ 0\le\theta_1\le 1 \qquad (16)$$

Similarly the marginal posterior density of θ_i (i=2,...,m-1) can be written from the (joint) posterior density (6). For the case of number of treatments m=2, taking $\theta=\theta_1$ the marginal posterior density of the parameter θ can easily be derived from (16) in the following form:

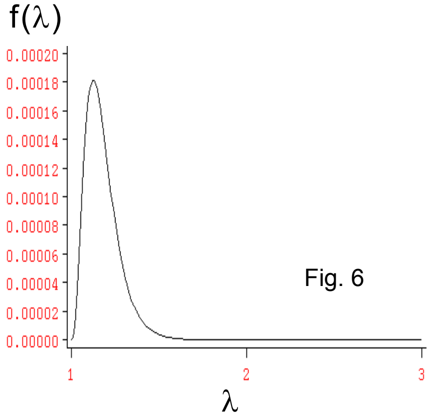
$$p(\theta|\mathbf{x}) \propto \theta^{a_1+n_{12}-1}(1-\theta)^{a_2+n_{21}-1}\int_1^{\infty}\frac{e^{-b_2(\lambda-1)}(\lambda-1)^{b_1+n_{o.12}-1}(\lambda+1)^{n_{o.12}}d\lambda}{\{\theta+\lambda(1-\theta)\}^{n_{12}}\{\lambda\theta+(1-\theta)\}^{n_{21}}},$$

$$0\le\theta\le 1. \quad (17)$$

The problem in obtaining an analytical expression of the integral in (17) is the same as in the integral of the marginal posterior density of λ though here the situation is more complicated. The following attempts are made for the approximate analytical solution of the integral in (17).

APPROXIMATION (I)

The shape of true integrand {say $f(\lambda)$} in (17) is examined (graph is given in Fig. 6), which is near to the Beta distribution. We can approximate it with a Beta distribution. The graph of the marginal posterior density of θ (17) is given in Fig. 7.



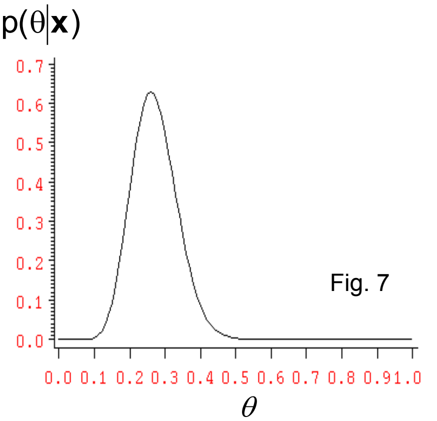


Fig. 6: Integrand of posterior (Marginal) density of θ i.e. $f(\lambda)$.

Fig. 7: Posterior (Marginal) density of θ i.e. $p(\theta|\mathbf{x})$.

APPROXIMATION (II)

The evaluation of the integral at different values of θ seems to be constant. So the integral can be absorbed in the constant of proportionality. Now the posterior density of θ can be written as:

$$p(\theta|\mathbf{x}) \propto \theta^{a_1+n_1-1}(1-\theta)^{a_2+n_2-1}, \tag{18}$$

which is a kernel of the beta distribution.

APPROXIMATION (III)

The lower bound and the upper bound can be fixed on the integral by neglecting some terms from the denominator of the integrand. In this way, the approximate analytical solution is in the form of a beta function.

References

Berger, J.O. (**1985**) “Statistical Decision Theory and Bayesian Analysis”, 2nd ed., Springer-Verlag, New York.

Bradley, R.A. (**1953**) “Some statistical methods in taste testing and quality evaluation”, *Biometrics*, 9, 22-38.

Bradley, R.A. and Terry, M.B. (**1952**) “Rank analysis of incomplete block design: I the method of paired comparisons”, *Biometrika,* 39, 324-345.

David, H.A. (**1988**) “The Method of Paired Comparisons”, 2nd ed., Charles Griffin and Company, London.

Erdelyi, A. *et al.* (**1954**) “Tables of Integrals Transforms”, Vol. I and II, McGraw-Hill, New York.

Lindley, D.V. (**1980**) “Approximate Bayesian methods, In: J.M. Bernardo, M.H. De Groot, D.V. Lindley and A.M.F. Smith (Eds.), *Bayesian Statistics*, University Press, Valencia, Spain, pp. 223-245.

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

Terry, L. and Kadane, J.B. (**1986**) "Accurate approximation for posterior moments and marginal densities”, *JASA*, 87, 82-86.