Bayesian estimation of Thurstonian ranking models based on the Gibbs sampler

Grace Yao*

Department of Psychology, National Taiwan University, Taiwan

Ulf Böckenholt

Department of Psychology, University of Illinois at Urbana-Champaign, USA

This paper presents a Gibbs sampler for the estimation of Thurstonian ranking models. This approach is useful for the analysis of ranking data with a large number of options. Approaches for assessing the goodness-of-fit of Thurstonian ranking models based on posterior predictive distributions are also discussed. Two simulation studies and two ranking studies are presented to illustrate that the Gibbs sampler is a promising solution to the numerical problems that previously plagued the estimation of Thurstonian ranking models.

1. Introduction

Few data collection techniques make weaker assumptions about the judgmental process than the ranking method. Only the ordinal information in the responses is utilized and, in contrast to the rating task, nuisance effects as a result of individual differences in response scale usage are avoided. Although there are a variety of approaches to choose from for the analysis of ranking data, a Thurstonian framework seems particularly well suited. Based on the assumption that ranking data are as a result of judgments following a multivariate normal distribution, Thurstonian models facilitate a straightforward formulation of both linear models for the means and structural equations for the covariance matrix of the judgments (Arbuckle & Nugent, 1973; Böckenholt, 1992, 1993).

Other approaches for analysing ranking data (e.g., Babington-Smith, 1950; Feigin & Cohen, 1978; Luce, 1959; Mallows, 1957; Pendergrass & Bradley, 1960) do not provide such a rich structure because they assume that options are judged independently of each other. If this assumption is violated, these models may be less informative for describing a data set. Both Marden (1995) and Fligner & Verducci (1993) provide comprehensive overviews of these and other ranking models proposed in the psychological and statistical literature. However, despite their obvious appeal, Thurstonian models are used only infrequently in applied research. The main reason for this lack of applications is related to difficulties in estimating the parameters of Thurstonian models. High-dimensional integration is required when analysing ranking data based on a large number of options. Unfortunately, numerical

* Requests for reprints should be addressed to Grace Yao, Department of Psychology, National Taiwan University, Taipei 10764, Taiwan.

integration and analytical approximation techniques are only useful when ranking a small number of options.

The major purpose of this paper is to present a computationally attractive approach based on the Gibbs sampler for estimating the parameters of Thurstonian ranking models. This method is especially useful for ranking problems with a large number of options. Special cases ranging from first choices, paired comparison and triad ordering, to full-rank orders can also be treated in the same framework. In fact, the approach presented is similar to that of McCulloch & Rossi (1994) who introduced the Gibbs sampling approach for the analysis of first-choice data with a multinomial probit model. Using a variant of the Gibbs sampler, these authors showed that the time-consuming evaluation of the likelihood function can be avoided by sampling from the exact posterior distribution of the multinomial probit model (see also Albert & Chib, 1993).

In addition to model estimation, it is shown how the Gibbs sampling approach can be used to assess the fit of high-dimensional Thurstonian ranking models by the construction and calculation of reference distributions for specific parameter-dependent statistics (Gelman, Meng & Stern, 1996). The practicability of the estimation and model assessment approach is demonstrated by the analysis of ranking data sets with four and ten options.

2. Thurstonian ranking models

In a ranking task, n judges are asked to rank m distinct options with respect to some criterion. Thurstone (1927) postulated that the ranking outcome is determined by latent random variables z_{ij} ($i = 1, ..., n; j = 1, ..., m$). The rank of option j for person i depends on the value of z_{ij}. For example, an option is ranked first if its value is largest and it is ranked second if its value is second largest, etc. Individual differences in the perception of the options are represented by the assumption that the z_{ij}s follow a multivariate normal distribution function with

$$z_{ij} = \beta_j^* + \varepsilon_{ij},$$

where β_j^* is the mean value of option j, and the joint distribution of ε_j is m-variate normal with mean vector $\mathbf{0}$ and covariance matrix Σ^*.

The ranking outcome is summarized by the ordering vector $\mathbf{r}$. For instance, when $\mathbf{r}_i = (h, j, ..., l, k)$, option h is judged superior to option j which in turn is judged superior to the remaining options, with the least preferred option being k. The probability of observing this rank order is

$$P(\mathbf{r}_i|\beta^*, \Sigma^*) = P((z_{ih} - z_{ij} > 0) \cap ... \cap (z_{il} - z_{ik} > 0)) = P(\mathbf{C}_i^* \mathbf{z}_i > \mathbf{0}),$$

where $\mathbf{C}_i^*$ is an $(m-1) \times m$ contrast matrix that describes the differences among the options. Because the $\mathbf{z}_i$s follow a multivariate normal distribution, the distribution of their differences is also multivariate normal. Consequently, the probability of a rank order can be determined by evaluating an $(m-1)$-variate normal distribution,

$$P(\mathbf{r}_i|\beta^*, \Sigma^*) = \frac{|\Gamma_i|^{-\frac{1}{2}}}{(2\pi)(m-1)/2} \int_0^\infty ... \int_0^\infty \exp\left\{-\tfrac{1}{2}(\delta_i - \mathbf{x})'\Gamma_i^{-1}(\delta_i - \mathbf{x})\right\}d\mathbf{x}, \tag{1}$$

where $\delta_i = \mathbf{C}_i^*\beta^*$ and $\Gamma_i = \mathbf{C}_i^*\Sigma^*(\mathbf{C}_i^*)'$.

Because of the difference structure, adding a constant to the β^* values does not change the ranking probabilities. As a result, a more compact representation is obtained by setting $\beta_m^* = 0$, or equivalently, by defining β_i $(i = 1, ..., m - 1)$ as $\beta_i^* - \beta_m^*$. In a similar fashion, we adjust the covariance matrix Σ^* by pre- and postmultiplying it by the contrast matrix $\mathbf{G}$ to obtain $\Sigma = \mathbf{G}\Sigma^*\mathbf{G}'$, where $\mathbf{G}$ is defined as

$$\mathbf{G} = \begin{bmatrix} 1 & 0 & 0 & 0 & \ldots & \ldots & -1 \\ 0 & 1 & 0 & 0 & \ldots & \ldots & -1 \\ \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & -1 \\ 0 & 0 & 0 & 0 & \ldots & 1 & -1 \end{bmatrix}.$$

Thus, δ_i and Γ_i can be rewritten as $\delta_i = \mathbf{C}_i\beta$ and $\Gamma_i = \mathbf{C}_i\Sigma(\mathbf{C_i})'$, where $\mathbf{C}_i$ is obtained from $\mathbf{C}_i^*$ by eliminating the mth column of $\mathbf{C}_i^*$.

When the number of options gets large, numerical integration of equation (1) becomes time-consuming. It is therefore not surprising that in applications of Thurstonian ranking models, only ranking probabilities with up to four options have been estimated. Clearly, there is a need for a method that facilitates the estimation of Thurstonian models for a larger number of options. The following section presents the Gibbs sampler as a feasible alternative to numerical integration.

3. The implementation of the Gibbs sampler

3.1. The Gibbs sampler

The Gibbs sampler, which was adapted from the Metropolis algorithm (Metropolis *et al.*, 1953; Hastings, 1970), was formally introduced by Geman & Geman (1984) in the context of image processing. The iterative Markov chain Monte Carlo (MCMC) algorithm facilitates the determination of joint posterior densities of many random variables by the simulation of various fully conditional distributions (for a detailed derivation and explanation, see Tanner, 1996; Chib & Greenberg, 1995). It has been shown that after a reasonably large number of iterations the joint density of the simulated random variables geometrically converges in distribution to the joint density of the true parameters. As a result, the Gibbs sampler provides a solution to a high-dimensional integration problem by replacing it with a sequence of easier calculations of conditional distributions. Below we apply this approach to the parameter estimation of Thurstonian ranking models. In this context, we find the Gibbs sampler to be particularly attractive for its conceptual and computational simplicity and ease of implementation.

3.2. Parameter estimation of Thurstonian ranking models

Given the matrix of observed ranking patterns $\mathbf{R}$ (with row vectors $\mathbf{r}_i$ $(i = 1, ..., n)$), we need to estimate the joint posterior distribution of (β, Σ)

$$\pi(\beta, \Sigma|\mathbf{R}) \propto \pi(\beta, \Sigma) \prod_{i=1}^{n} \mathrm{P}(\mathbf{r}_i|\beta, \Sigma).$$

Following the approach by McCulloch & Rossi (1994), we employ a normal prior $N(\beta_0, \Sigma_0)$

for estimating β, i.e.,

$$\pi(\beta|\beta_0, \Sigma_0) \propto |\Sigma_0|^{-\frac{1}{2}} \exp\{-\tfrac{1}{2}(\beta - \beta_0)'\Sigma_0^{-1}(\beta - \beta_0)\}$$

and an independent Wishart prior, $W(\upsilon_0, \mathbf{V}_0)$, for Σ_0^{-1}, i.e.,

$$\pi(\Sigma^{-1}|\upsilon_0, \mathbf{V}_0) \propto |\Sigma^{-1}|(\upsilon_0 - \mathrm{m})/2 \exp\left[\mathrm{tr}\left\{-\frac{1}{2}\Sigma^{-1}\mathbf{V}_0^{-1}\right\}\right].$$

To minimize the influence of the priors, they are specified to be diffuse. Thus, υ_0 is set to a small number and the variances of β_0 are specified to be large. In the reported applications, we used $\upsilon_0 = \mu + 1, \Sigma_0 = 100\,\mathbf{G}'\mathbf{G}, \beta_0 = \mathbf{0}$, and specified $\mathbf{V}_0 = \upsilon_0\mathbf{G}'\mathbf{G}$ as part of the prior for the covariance matrix. The estimation of the posterior density $\pi[\beta, \Sigma|\mathbf{R}]$ is difficult because it requires the evaluation of the likelihood function. Fortunately, this problem can be avoided by the introduction of the latent judgments $[z_{ij}]$. When augmenting (β, Σ) by the $n \times (m - l)$ matrix $\mathbf{Z}$ with elements $[z_{ij} - z_{im}]$, the joint posterior density of $[\mathbf{Z}, \beta, \Sigma]$ given $\mathbf{R}$ is

$$\pi(\mathbf{Z}, \beta, \Sigma|\mathbf{R}) \propto \pi(\mathbf{R}|\mathbf{Z})\pi(\mathbf{Z}|\beta, \Sigma)\pi(\beta, \Sigma).$$

Based on this decomposition, it is straightforward to specify the Gibbs sampler via the following sequence of conditional distributions of $\mathbf{z}_i$, β and Σ : $\pi(\mathbf{z}_i|\mathbf{r}_i, \beta, \Sigma)$ for $i = 1, 2, ..., n$ (step 1); $\pi(\beta|\mathbf{Z}, \Sigma)$ (step 2); $\pi(\Sigma|\mathbf{Z}, \beta)$ (step 3). Clearly, the information provided by $\mathbf{R}$ is redundant when $\mathbf{Z}$ is available. Next, we discuss each of the steps in detail.

Step 1. Given β, Σ and $\mathbf{r}_i$, one draws from the conditional posterior of $\mathbf{x}_i\,(i = 1, 2, ..., n)$, which is an $(m - 1)$-variate normal distribution truncated at $\mathbf{0}$ with mean $\mathbf{C}_i\beta$ and covariance matrix $\mathbf{C}_i\Sigma\mathbf{C}_i'$. The latent judgment vector z_i is obtained from $\mathbf{x}_i$ by pre-multiplying $\mathbf{x}_i$ by the inverse of $\mathbf{C}_i, \mathbf{z}_i = \mathbf{C}_i^{-1}\mathbf{x}_i$.

The truncated multivariate normal distribution of $\mathbf{x}_i$ is readily simulated by Hajivassiliou's (1993) truncated multivariate normal (TMVN) Gibbs sampler. This method cycles repeatedly through $m - 1$ fully conditional, truncated univariate normal distributions and converges to the TMVN as a limiting distribution. A detailed discussion of this approach is provided by Hajivassiliou, McFadden & Ruud (1996); FORTRAN and GAUSS computer code of the TMVN Gibbs sampler are available via anonymous ftp from econ.yale.edu. Vijverberg (1997) reviews several comparative studies which demonstrate that this Gibbs sampler simulator is among the best of currently available simulation algorithms for the (truncated) multivariate normal distribution.

Step 2. Draws from the conditional distribution of $[\beta|\mathbf{Z}, \Sigma]$ are obtained from a normal distribution since

$$\beta|\mathbf{Z}, \Sigma \sim N(\beta^*, \Sigma^*),$$

where

$$\Sigma^* = (\Sigma_0^{-1} + n\Sigma^{-1})^{-1}$$

and

$$\beta^* = \Sigma^*(\Sigma_0^{-1}\beta_0 + n\Sigma^{-1}\bar{\mathbf{z}}),$$

in which $\bar{\mathbf{z}}$ contains the column means of $\mathbf{Z}$ (see Gelman, Carlin, Stern & Rubin, 1995).
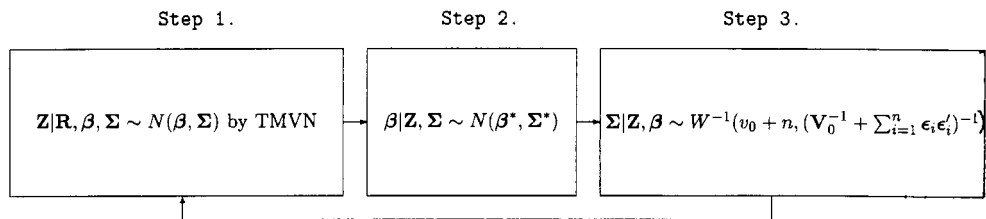
Step 1.	Step 2.	Step 3.
$\mathbf{Z}|\mathbf{R}, \beta, \Sigma \sim N(\beta, \Sigma)$ by TMVN	$\beta|\mathbf{Z}, \Sigma \sim N(\beta^*, \Sigma^*)$	$\Sigma|\mathbf{Z}, \beta \sim W^{-1}(v_0 + n, (\mathbf{V}_0^{-1} + \sum_{i=1}^{n} \epsilon_i \epsilon_i')^{-1})$

Figure 1. Markov chain Monte Carlo for the estimation of Thurstonian ranking models.

Step 3. The conditional distribution of $\Sigma^{-1}|\mathbf{Z}, \beta$ can be derived as a Wishart posterior by combining the conjugate Wishart prior with the likelihood

$$[\Sigma^{-1}|\mathbf{Z}, \beta] \sim W\left(v_0 + n, \left(\mathbf{V}_0^{-1} + \sum_{i=1}^{n} \varepsilon_i \varepsilon_i'\right)^{-1}\right),$$

where

$$\varepsilon_i = \mathbf{z}_i - \beta.$$

Thus, the posterior distribution of $\Sigma|\mathbf{Z}, \beta$ is inverse Wishart, denoted by $W^{-1}(\cdot)$ (see Zellner, 1971). An efficient method for obtaining draws from this conditional distribution is first to generate the precision matrix Σ^{-1} from the posterior Wishart distribution and then to invert it. Samples from the Wishart distribution are generated by using Bartlett's (1933) decomposition (see also Johnson, 1987).

The flow diagram in Fig. 1 summarizes the MCMC for the estimation of Thurstonian ranking models. After repeating the draws from the three conditional distribution many times, the draws will converge to a single draw from the joint posterior of $\mathbf{Z}$, β and Σ. The sampling process is continued to obtain additional draws from the posterior.

Predictions of Thurstonian ranking models are invariant under elementwise, monotonic and strictly increasing transformations such as positive scalar multiplication and scalar shifts. Our approach to this identification problem is to report the identified parameters given by the standardized mean differences $\beta_j / \sqrt{\sigma_{jj}}$, the standardized variance ratios defined as σ_{jj}/σ_{11}, and the correlations of the mean differences defined as $\sigma_{jk} / \sqrt{\sigma_{jj}\sigma_{kk}}$ ($j, k = 1, ..., m - 1$). To simplify the interpretation of the identified parameters, it is convenient to compute Σ^* from Σ. A unique solution can be obtained by setting the variances of Σ^* to 1 and constraining one of the covariances to a fixed value (Dansie, 1986). This approach is illustrated in the application section.

Computations were carried out in Version 3 of GAUSS (1995). A listing of the GAUSS program is available upon request from the authors. Our implementation of the Gibbs sampler on a 486 PC with 16MB RAM took about 5 seconds per cycle with 10 options and 178 distinct rankings and about 0.3 seconds per cycle with four options and 24 distinct rankings.

3.3. Convergence diagnostics

Because, in general, it is unclear how many iterations are necessary until convergence, it is important to investigate this issue in each application. Although a variety of convergence

diagnostics are available for this purpose, it should be noted that it is not possible to determine with certainty whether a finite Gibbs sample is representative of an underlying stationary distribution (see Cowles & Carlin, 1996, for a recent review of convergence diagnostics).

To assess convergence in each of our reported applications, we ran at least three parallel chains with a minimum of 11 000 draws. In addition to visually inspecting the chains by overlaying their sampled values on a common graph, we computed autocorrelations and several convergence diagnostics proposed by Geweke (1992), Gelman & Rubin (1992; see also Brooks & Gelman, 1998), Raftery & Lewis (1992) and Heidelberger & Welch (1983). These diagnostics are implemented in a freely available S-PLUS program called CODA (Best, Cowles & Vines, 1997).

In the most complex application with ten options and 53 parameters to be estimated (n = 178) convergence occurred in less than 2000 draws. In all other cases, convergence to the stationary distribution occurred more rapidly with a considerably smaller number of iterations.

4. A simulation study

In this section, we present the results of two small simulation studies that compare results obtained by maximum likelihood estimation and the Gibbs sampler for two different sample sizes. Data sets were generated for m = 4 options and n = 500 or n = 100 judges in the first and the second study, respectively, with mean vector

$$\beta^* = [-.25, .25, 0, -.25]$$

and covariance matrix

$$\Sigma^* = \begin{bmatrix} 1 & .6 & .3 & .1 \\ .6 & 1 & .6 & .3 \\ .3 & .6 & 1 & .6 \\ .1 & .3 & .6 & 1 \end{bmatrix}.$$

There are $m(m + 1)/2 - 2$ identified parameters to be determined. The identified parameters include three standardized means, two variance ratios, and three correlations of the mean differences.

For both sample sizes, 1000 data sets were generated and fitted by maximum likelihood methods. The third and sixth columns of Table 1 contain the means of the parameter estimates and their standard deviation for n = 500 and n = 100, respectively. We note that for n = 500 all parameter estimates are virtually unbiased but that for n = 100, the variance ratios are overestimated and the correlations of mean differences are underestimated. The remaining columns contain maximum likelihood estimates and the posterior means and standard deviations (in parentheses) of the Gibbs sampler for a single generated data set. Posterior means and maximum likelihood estimates agree to two decimal places for n = 500. For n = 100, differences between posterior means and maximum likelihood estimates are somewhat larger but still very minor with respect to the variance ratios and correlations of mean differences. The posterior standard deviations and the standard deviations of the maximum likelihood estimates are also very similar. Clearly, these results indicate that the Gibbs sampler provides a promising approach for the estimation of Thurstonian ranking models.

Table 1. Identified population parameter values, maximum likelihood (based on 1000 and 1 replications, respectively) and Gibbs sampler estimates (based on 1 replication)

Parameter	True	$n = 500$			$n = 100$		
		ML (1000)	ML (1)	Gibbs (1)	ML (1000)	ML (1)	Gibbs (1)
$\dfrac{\beta_1}{\sqrt{\sigma_{11}}}$	.000	.000 (.051)	−.028	−.029 (.052)	.000 (.116)	.173	.176 (.117)
$\dfrac{\beta_2}{\sqrt{\sigma_{22}}}$	.423	.425 (.054)	.424	.424 (.054)	.425 (.123)	.587	.591 (.123)
$\dfrac{\beta_3}{\sqrt{\sigma_{33}}}$	.280	.280 (.054)	.348	.346 (.055)	.281 (.120)	.600	.595 (.127)
$\dfrac{\sigma_{22}}{\sigma_{11}}$	.778	.779 (.071)	.799	.804 (.079)	.792 (.178)	.927	.954 (.197)
$\dfrac{\sigma_{33}}{\sigma_{11}}$	.444	.447 (.056)	.417	.425 (.055)	.453 (.126)	.425	.454 (.120)
$\dfrac{\sigma_{12}}{\sqrt{\sigma_{11}\sigma_{22}}}$	.756	.755 (.033)	.720	.718 (.034)	.756 (.081)	.768	.756 (.067)
$\dfrac{\sigma_{13}}{\sqrt{\sigma_{11}\sigma_{33}}}$	.500	.498 (.054)	.460	.459 (.056)	.493 (.120)	.558	.549 (.112)
$\dfrac{\sigma_{23}}{\sqrt{\sigma_{22}\sigma_{33}}}$	.661	.660 (.042)	.601	.607 (.044)	.652 (.100)	.683	.663 (.096)

Note: Figures in parentheses are standard deviations in the case maximum likelihood estimates posterior standard deviations in the case of Gibbs sampler estimates.

4.1. Model fit tests: The posterior predictive check method

Typically, the fit of a ranking model is assessed by computing observed and expected frequencies of ranking patterns based on the Pearson or likelihood ratio statistics. This approach is not feasible when the number of options to be ranked is large because only a small subset of the possible ranking patterns may be observed. As a result, there is little justification that the likelihood ratio statistic will follow a chi-square distribution under the null hypothesis. Instead, model fit may be assessed by comparing standardized differences between the observed and predicted ranking probabilities in different partitions of the data (Cohen & Mallows, 1983). Although the distribution of a fit statistic comparing certain partitions of the ranking data is not known, it can easily be generated within the Gibbs sampling framework. In Section 5, we apply the posterior predictive check (PPC) method for model assessment. The use of posterior predictive checks to assess model fitness was introduced by Guttman (1967). Recent applications and extensions of this work can be found in Gelman *et al.* (1996).

Suppose the Gibbs sampler converged after $t - 1$ iterations. Starting from the tth iteration, H (possibly dependent) draws are collected. A data set containing n ranking patterns, $\mathbf{R}^{\text{rep}_h}$ ($h = 1, ..., H$) is simulated for each of the H parameter vectors. Next, some statistic is calculated for each of the H generated data sets and tabulated. The resulting distribution is called reference distribution because it is used to compute the percentile rank of the statistic obtained from the observed data. Statistics with low or high percentile ranks are indicative of misfit of the model. Although this approach is useful in assessing local (mis)fits as defined by the statistic, with a slight modification, it can also be applied for an overall goodness-of-fit assessment (Gelman *et al.*, 1996).

To assess the fit of a Thurstonian ranking model, we compute three discrepancy measures

based on the paired comparisons, triple and quadruple rankings in a data set. Inspecting the fit of higher-order rankings is also possible but may be limited by the sample size n. For the paired comparison discrepancy measure, we compute for each generated data set and the observed data the number of times option i is preferred to option j ($i \neq j; i, j = 1, ..., m$). These paired comparison probabilities for the generated and observed data sets are referred to as $\pi_{ij}^{\text{rep}_h}$ and π_{ij}^{obs}. In addition, we compute the corresponding expected paired comparison probabilities

$$\hat{\pi}_{lj}^{(h)} = \Phi\left(\frac{\hat{\beta}_l^{(h)} - \hat{\beta}_j^{(h)}}{\sqrt{(\hat{\sigma}_l^{(h)})^2 + (\hat{\sigma}_j^{(h)})^2 - 2\hat{\sigma}_{lj}^{(h)}}}\right),$$

and a discrepancy measure for the generated data

$$T^{\text{rep}_h}(\mathbf{R}^{\text{rep}_h}, \beta^{\text{rep}_h}, \Sigma^{\text{rep}_h}) = n \sum_{l<j} \frac{(\pi_{lj}^{\text{rep}_h} - \hat{\pi}_{lj}^{(h)})^2}{\hat{\pi}_{lj}^{(h)}},$$

as well as for the observed data

$$T^{\text{obs}}(\mathbf{R}^{\text{obs}}, \beta^{\text{rep}_h}, \Sigma^{\text{rep}_h}) = n \sum_{l<j} \frac{(\pi_{lj}^{\text{obs}} - \hat{\pi}_{lj}^{(h)})^2}{\hat{\pi}_{lj}^{(h)}}.$$

Finally, the posterior predictive p-value, PPP, for the paired comparison is determined as

$$PPP = \text{P}[T^{\text{rep}_h}(\mathbf{R}^{\text{rep}_h}, \beta^{\text{rep}_h}, \Sigma^{\text{rep}_h}) \geq T^{\text{obs}}(\mathbf{R}^{\text{obs}_h}, \beta^{\text{rep}_h}, \Sigma^{\text{rep}_h})].$$

The same procedure is followed for the triple and quadruple rankings that can be obtained from the observed and generated data sets.

5. Applications

5.1. Desirability of political goals

To illustrate the PPC method, we selected Croon's (1989) ranking data of 2262 German respondents about the desirability of the four political goals: (1) the maintenance of order in the nation; (2) giving people more say in the decisions of government; (3) fighting rising prices; (4) protecting freedom of speech. In his analysis of the data set, Böckenholt (1993) showed that the Thurstonian model is not well suited to describe the rankings. Instead, he found that a ranking model allowing for two distinct subpopulations provided a more adequate representation of the data: One subpopulation preferred items (1) and (3), while the other subgroup preferred items (2) and (4). This result is consistent with Inglehart's (1977) distinction between a materialistic ((1), (3)) and post-materialistic ((2), (4)) value orientation in political goals.

Table 2 contains the identified parameter estimates obtained by maximum likelihood and the Gibbs sampler. As in the simulation study, there is almost perfect agreement between the estimates. To simplify the interpretation of the covariance structure, we set $\sigma_{34}^* = 0$ and $\sigma_{ii}^* = 1 (i = 1, ..., 4)$. The resulting covariance matrix is given is Table 3(a). Note that items (1), (3) and items (2), (4) are more highly correlated with each other than with any of the other items. The selection of the fixed value for σ_{34}^* is arbitrary as long as Σ^*

Table 2. Parameter estimates obtained by maximum likelihood and Gibbs sampler for Croon (1989) data

Parameter	ML	Gibbs
$\dfrac{\beta_1}{\sqrt{\sigma_{11}}}$	.727	.727 (.028)
$\dfrac{\beta_2}{\sqrt{\sigma_{22}}}$	.162	.162 (.025)
$\dfrac{\beta_3}{\sqrt{\sigma_{33}}}$	.798	.798 (.028)
$\dfrac{\sigma_{22}}{\sigma_{11}}$	.583	.587 (.042)
$\dfrac{\sigma_{33}}{\sigma_{11}}$	.926	.927 (.042)
$\dfrac{\sigma_{12}}{\sqrt{\sigma_{11}\sigma_{22}}}$	.336	.336 (.028)
$\dfrac{\sigma_{13}}{\sqrt{\sigma_{11}\sigma_{33}}}$	.675	.674 (.021)
$\dfrac{\sigma_{23}}{\sqrt{\sigma_{22}\sigma_{33}}}$	.393	.393 (.028)

Note: figures in parentheses are the posterior standard deviations.

remains positive definite. In Tables 3(b) and 3(c), we depict two equivalent solutions for Σ^* by setting σ^*_{34} equal to .3 and .6, respectively. Note that the order of the elements of Σ^* is not affected by the choice of value for σ^*_{34}. It is straightforward to show that the relationship between the elements of two different covariance matrices $\Sigma^{*(1)}$ and $\Sigma^{*(2)}$ is linear and given by

$$\sigma^{*(1)}_{lj} = (1 - \phi) + \phi\,\sigma^{*(2)}_{lj}$$

Table 3. Three equivalent covariance matrices Σ^* for the Croon (1989) data

(a)			
1.000			
−.156	1.000		
.323	−.007	1.000	
−.080	.370	.000*	1.000
(b)			
1.000			
.191	1.000		
.526	.295	1.000	
.244	.559	.300*	1.000
(c)			
1.000			
.538	1.000		
.729	.597	1.000	
.568	.748	.600*	1.000

Note: Values superscripted with * are fixed.

where

$$\phi = \frac{1 - \frac{1}{m^2}\Sigma_a\Sigma_b\sigma_{ab}^{*(1)}}{1 - \frac{1}{m^2}\Sigma_a\Sigma_b\sigma_{ab}^{*(2)}}.$$

Pearson's χ^2 test of the observed and expected ranking probabilities estimated by maximum likelihood, $P^2 = 60.8\,(df = 23 - 8 = 15, p < .001)$, indicates that the Thurstonian model does not provide an adequate representation of the data. A similar but more detailed result is provided by the PPC method. The estimated posterior predictive p-values (based on $H = 500$ draws) are .204 for the paired comparison margins, .004 for the triple rankings and .000 for the quadruple rankings. Thus, only the observed paired comparison probabilities are fitted well by the Thurstonian ranking model.

This example demonstrates that the discrepancy measures can be useful in determining overall model fit. In particular, when the sample size is small in comparison to the number of possible ranking outcomes, the PPC method is an important tool in detecting systematic misfit.

5.2. Preference for gifts

The following data set consists of the rankings of 10 options from a study on compound preferences (McKeon, 1961). McKeon was interested in determining whether a preference value of compound stimuli consisted of two options can be predicted by the additive combination of the preference values of each of the two options. To test this hypothesis, he applied a partially balanced incomplete block design by constructing 11 blocks with 10 single and/or composite alternatives in each block. One hundred and seventy-eight subjects were instructed to rank the ten (either single and/or compound) alternatives within each block based on their suitability as gifts for friends. The first block contained ten single gifts. Each of the other blocks contained one single item and nine pairs of items as compound stimuli. The ten possible single gifts were: (1) camera; (2) typewriter; (3) portable radio; (4) record player (5) dictionary; (6) pen and pencil set; (7) cigarette case and lighter; (8) briefcase; (9) bookcase; and (10) desk lamp.

McKeon (1961) converted the ranking results for each block to paired comparison data which he used to estimate the scale values under the Case V constraint (i.e., $\Sigma^* = \mathbf{I}$). In our analysis of the McKeon data, we focus on the first block that contains the rankings of the ten single options. Table 4 gives the 'option by rank position' table based on the rankings of the 178 subjects. The last two columns are the mean ranks and the corresponding standard deviations, respectively.

To compare our results with those obtained by McKeon (1961) we also fitted the Case V model to the data. Table 5 contains the standardized scale values obtained by McKeon (1961) in the second column and the posterior means obtained by the Gibbs sampler in the third column. The corresponding posterior standard deviations are given in parentheses. The correlation between the scale values of McKeon and the posterior means is very high (.998), despite the fact that the Gibbs sampler considers the full information in the rankings while McKeon obtained his results from the paired comparison data under the wrong assumption that the pairs are independent. However, the Case V model does not provide an adequate representation of the data. The posterior predictive p-values (based on $H = 500$ draws) are

Table 4. ''Object by Ranking Position'' Table for McKeon's (1961) data after rank transformation (rank 1, most preferred; 10, least preferred)

Object	Rank position										Mean rank	Std. err.
	1	2	3	4	5	6	7	8	9	10		
O_1	37	30	33	35	10	9	6	8	5	5	3.60	.18
O_2	45	49	34	24	7	7	6	1	2	3	2.90	.15
O_3	35	42	36	31	13	6	6	5	3	1	3.21	.15
O_4	33	29	24	23	19	16	8	15	10	1	4.06	.19
O_5	4	9	13	13	20	28	30	27	26	8	6.30	.17
O_6	2	4	3	10	29	31	29	29	34	7	6.75	.15
O_7	3	3	2	2	9	12	11	17	22	97	8.55	.16
O_8	4	1	7	5	7	18	16	29	49	42	7.84	.16
O_9	13	7	14	23	42	17	27	18	12	5	5.42	.17
O_{10}	2	4	12	12	22	34	39	29	15	9	6.39	.15

O_1, camera; O_2, typewriter; O_3, portable radio; O_4, record player; O_5, dictionary; O_6, pen and pencil set; O_7, cigarette case and lighter; O_8, briefcase; O_9, bookcase; O_{10}, desk lamp.

both .000 for the paired comparison margins and triple rankings. (Quadruple rankings were not computed because of the small sample size.) In contrast, the Thurstonian model with an unconstrained covariance matrix yields a much better fit of the data. The posterior predictive p-values (based on $H = 500$ draws) are .518 for the paired comparison margins and .270 for the triple rankings. The fourth column of Table 5 contains the estimated standardized posterior means and standard deviations of the general Thurstonian model. Interestingly, the

Table 5. McKeon's (1961) Case V solution and mean vector of estimated posterior parameters (with standard deviation) for Case V model and unconstrained covariance matrix ($m = 10, n = 178$)

	McKeon's scale values	$\Sigma^* = \mathbf{I}$	Σ^* general
$\dfrac{\hat{\beta}_1}{\sqrt{\hat{\sigma}_{11}}}$	.845	.920 (.084)	.841 (.093)
$\dfrac{\hat{\beta}_2}{\sqrt{\hat{\sigma}_{22}}}$	1.092	1.141 (.084)	1.180 (.108)
$\dfrac{\hat{\beta}_3}{\sqrt{\hat{\sigma}_{33}}}$	1.047	1.037 (.083)	1.030 (.101)
$\dfrac{\hat{\beta}_4}{\sqrt{\hat{\sigma}_{44}}}$	.699	.792 (.082)	.743 (.089)
$\dfrac{\hat{\beta}_5}{\sqrt{\hat{\sigma}_{55}}}$	.030	.038 (.082)	.034 (.083)
$\dfrac{\hat{\beta}_6}{\sqrt{\hat{\sigma}_{66}}}$	−.109	−.097 (.078)	−.132 (.083)
$\dfrac{\hat{\beta}_7}{\sqrt{\hat{\sigma}_{77}}}$	−.725	−.836 (.085)	−.831 (.093)
$\dfrac{\hat{\beta}_8}{\sqrt{\hat{\sigma}_{88}}}$	−.462	−.498 (.084)	−.626 (.089)
$\dfrac{\hat{\beta}_9}{\sqrt{\hat{\sigma}_{99}}}$	.292	.310 (.080)	.387 (.084)

Table 6. Posterior means and standard deviations for Σ for the McKeon (1961) data (values above and below the diagonal of the first matrix are covariances and correlations, respectively)

Means								
1.000	.601	.652	.664	.208	.236	.202	.125	.287
.705	.727	.592	.572	.180	.122	.151	.081	.178
.718	.764	.825	.714	.119	.178	.238	.044	.218
.650	.657	.770	1.043	.125	.175	.235	.094	.206
.327	.333	.207	.193	.403	.133	.134	.125	.155
.429	.260	.357	.312	.382	.302	.179	.090	.104
.214	.188	.278	.244	.223	.346	.890	.203	.124
.191	.144	.073	.140	.299	.249	.327	.431	.094
.431	.212	.360	.303	.368	.284	.198	.216	.443

Std. Dev.								
fixed								
.053	.099							
.051	.045	.108						
.059	.060	.043	.141					
.082	.087	.089	.086	.076				
.076	.090	.082	.081	.081	.060			
.088	.093	.087	.089	.092	.087	.208		
.089	.094	.094	.089	.087	.091	.088	.096	
.078	.088	.085	.086	.081	.085	.090	.091	.080

standardized scale values are very similar to the Case V solutions. The posterior means and standard deviations of the Σ variates are given in Table 6. We note that the gifts can be separated into two groups consisting of options 1 to 4 and 5 to 10. This result suggests a simplified representation of the covariance matrix with equal correlations between the options within each group.

In summary, these results show that the general Thurstonian ranking model provides a satisfactory account of the data. McKeon's (1961) restrictive assumption that the gift options are perceived independently by the respondents is not supported. Instead, there appear to be systematic dependencies in the ranking data that are captured by the unconstrained covariance matrix.

6. Conclusion

This paper discussed an application of the Gibbs sampler for estimating the parameters of Thurstonian ranking models and applied a model check procedure based on the posterior distribution of the parameters. Two simulation studies and two applications demonstrated that the Gibbs sampler facilitates the application of Thurstonian ranking models to large-scale problems which encountered computational difficulties in the past.

Since the Gibbs sampler is still a relatively new technique, there is a need for further research and applications of this procedure. Open issues include well-defined methods for assessing convergence of the Gibbs sampler, a comprehensive list of Thurstonian models based on structural equations and their implementation via the Gibbs sampler, and more diagnostic checks and specification tests. However, despite this long list of future work, we

expect many more applications of the rich and flexible class of Thurstonian ranking models in the near future. The Markov chain Monte Carlo approach appears to be a promising solution to the numerical intractabilities that previously plagued the estimation of Thurstonian ranking models.

Acknowledgements

References

Albert, J. & Chib, S. (1993). Bayesian analysis of binary and polychotomous data. *Journal of the American Statistical Association*, **88,** 669–679.

Arbuckle, J. & Nugent, J. H. (1973). A general procedure for parameter estimation for the law of comparative judgment. *British Journal of Mathematical and Statistical Psychology*, **26,** 240–260.

Babington-Smith, B. (1950). Discussion of Professor Ross' paper. *Journal of the Royal Statistical Society B*, **12,** 153–162.

Bartlett, M. S. (1933). On the theory of statistical regression. *Proceedings of the Royal Society of Edinburgh*, **53,** 260–283.

Best, N., Cowles, M. K. & Vines, K. (1997). *CODA: Convergence diagnosis and output analysis software for Gibbs sampling output, Version 0.40 (Addendum to Manual)*. Technical report, MRC Biostatistics Unit, University of Cambridge.

Böckenholt, U. (1992). Thurstonian models for partial ranking data. *British Journal of Mathematical and Statistical Psychology*, **45,** 31–49.

Böckenholt, U. (1993). Applications of Thurstonian models to ranking data. In M. Fligner & J. Verducci (Eds), *Probability models and statistical analyses for ranking data*, pp. 157–172. New York: Springer-Verlag.

Brooks, S. P. & Gelman, A. (1998). General methods for monitoring convergence of iterative simulations. *Journal of Computational and Graphical Statistics*, **7**, 434–455.

Chib, S. & Greenberg, E. (1995). Understanding the Metropolis–Hastings algorithm. *American Statistician*, **49,** 327–335.

Cohen, A. & Mallows, C. L. (1983). Assessing goodness of fit of ranking models to data. *The Statistician*, **32,** 361–373.

Cowles, M. K. & Carlin, B. P. (1996). Markov chain Monte Carlo convergence diagnostics: A comparative review. *Journal of the American Statistical Association*, **91,** 883–904.

Croon, M. (1989). Latent class models for the analysis of rankings. In G. DeSoete, H. Feger & K. C. Klauer (Eds), *New developments in psychological choice modeling*, pp. 99–121. Amsterdam: Elsevier.

Dansie, B. R. (1986). Normal order statistics as permutation probability models. *Applied Statistics*, **35,** 269–275.

Feigin, P. D. & Cohen, A. (1978). On a model for concordance between judges. *Journal of the Royal Statistical Society B*, **40,** 203–213.

Fligner, M. & Verducci, J. (Eds) (1993). *Probability models and statistical analyses for ranking data*. New York: Springer-Verlag.

GAUSS (1995). *Systems and graphics manual. Version 3*. Kent: Aptech Systems Inc.

Gelman, A., Carlin, J. B., Stern, H. S. & Rubin, D. B. (1995). *Bayesian data analysis*. London: Chapman & Hall.

Gelman, A., Meng, X. L. & Stern, H. S. (1996). Posterior predictive assessment of model fitness via realized discrepancies. *Statistica Sinica*, **6,** 733–807.

Gelman, A. & Rubin, D. B. (1992). Inference from iterative simulation using multiple sequences (with discussion). *Statistical Science*, **7,** 457–511.

Geman, S. & Geman, D. (1984). Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **6,** 721–741.

Geweke, J. (1992). Evaluating the accuracy of sampling-based approaches to the calculation of posterior moments. In J. M. Bernardo, J. Berger, A. P. David & A. F. M. Smith (Eds), *Bayesian statistics 4*, pp. 169–193. Oxford: Oxford University Press.

Guttman, I. (1967). The use of the concept of a future observation in goodness-of-fit problems. *Journal of the Royal Statistical Society B*, **29,** 83–100.

Hajivassiliou, V. A. (1993). Simulation estimation methods for limited dependent variable models. In G. S. Maddala, C. R. Rao & H. D. Vinod (Eds), *Handbook of statistics*, vol. 11. Amsterdam: Elsevier.

Hajivassiliou, V. A., McFadden, D. & Ruud, P. (1996). Simulation of multivariate normal rectangle probabilities and their derivatives. Theoretical and computational results. *Journal of Econometrics*, **72,** 85–134.

Hastings, W. K. (1970). Monte Carlo sampling methods using Markov chains and their applications. *Biometrika*, **57,** 97–109.

Heidelberger, P. & Welch, P. D. (1983). Simulation run length control in the presence of an initial transient. *Operations Research*, **31,** 1109–1144.

Inglehart, R. (1977). *The silent revolution: Changing values and political styles among Western publics.* Princeton, NJ: Princeton University Press.

Johnson, M. E. (1987). *Multivariate statistical simulation.* New York: Wiley.

Luce, R. D. (1959). *Individual choice behaviour.* New York: Wiley.

Mallows, C. L. (1957). Non-null ranking models. I. *Biometrika*, **44,** 114–130.

Marden, J. I. (1995). *Analyzing and modeling rank data.* New York: Chapman & Hall.

McCulloch, R. & Rossi, P. E. (1994). An exact likelihood analysis of the multinomial probit model. *Journal of Econometrics*, **64,** 207–240.

McKeon, J. J. (1961). Measurement procedures based on comparative judgments. Unpublished doctoral dissertation. University of North Carolina, Chapel Hill.

Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H. & Teller, E. (1953). Equations of state calculations by fast computing machines. *Journal of Chemical Physics*, **21,** 1087–1091.

Pendergrass, P. N. & Bradley, R. A. (1960). Ranking in triple comparison. In I. Olkin (Ed.), *Contributions to probability and statistics*, pp. 331–351. Palo Alto, CA: Stanford University Press.

Raftery, A. E. & Lewis, S. (1992). How many iterations in the Gibbs sampler? In J. M. Bernardo, J. Berger, A. P. Dawid & A. F. M. Smith (Eds), *Bayesian statistics 4*, pp. 763–773. Oxford: Oxford University Press.

Tanner, M. A. (1996). *Tools for statistical inference*, 3rd ed. New York: Springer-Verlag.

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

Vijverberg, W. P. M. (1997). Monte Carlo evaluation of multivariate normal probabilities. *Journal of Econometrics*, **76,** 281–307.

Zellner, A. (1971). *Introduction to Bayesian inference in econometrics.* New York: Wiley.