Bayesian inference for the CRS model for ranking data

Rebecca Gordon[1,2], **Geoff Nicholls**[1]
1. Department of Statistics, University of Oxford
2. Heriot-Watt University
rg87@hw.ac.uk

Abstract

Rankings are a common data type and their inference is important in many fields. Context dependence and transitivity can be restricting model assumptions for this data type, and thus the CRS model has been proposed for situations where these assumptions are false. We present the first (equal) Bayesian inference of the CRS model, testing this on synthetic and empirical data. We also discuss some aspects of the CRS model in further detail.

1 Introduction

A ranking is an ordered lists of a set of "actors" such that the order in which the actors appear in the list denotes some kind of "preference". This could for example be the preference of a decision maker, e.g. in a ranking of food types, we may have (1 - Pizza, 2 - Ice Cream, 3 - Burgers), denoting that Pizza is the most preferred, Ice Cream second and Burgers third. The "preference" could also denote a finishing position in a race, as seen with the Formula 1 data used later in this paper. Usually, we wish to infer the unknown "true" ordering of the actors.

Ranking data occurs naturally in many fields such as consumer behaviour [Pennock et al. [2000]], medicine [Zhang et al. [2021]] and educational psychology [Bargagliotti et al. [2021]].

Multiple models have been proposed for ranking data. Many of these view a ranking as a "top-down" choice problem, i.e. that a ranking is generated by sequentially choosing an actor to go in each position, starting at the top. The probability of a given ranking is then the joint probability of all the choices made in its generation.

A popular model for choice analysis is the Multinomial Logit Model (MNL) [Luce [1959]] which assigns a "utility" to each actor, such that the higher an actors utility, the more preferable it is. However this model requires the satisfaction of Luce's Choice Axiom, and therefore allows no context dependence or intransitivity. The context-dependent utility model (CDM) [Seshadri et al. [2019]] is a generalisation of the MNL which assigns a parameter to each ordered pair of distinct actors. The probability of a choice is then constructed such that the "utility" of a given actor can change based on the other actors present in the set. This means that the CDM does not require the satisfaction of Luce's choice axiom, and can thus be used to model choice systems that contain context dependence and intransitivity.

The Plackett-Luce (PL) model [Plackett [1975]] is a model for rankings where each choice in the ranking is modelled by the MNL. Likewise, the Contextual Repeated Selection(CRS) model[Seshadri et al. [2020]] is a model for rankings where each choice in the ranking is modelled by the CDM. Just as the CDM is a generalisation of the MNL, the CRS model is a generalisation of PL in which Luce's choice axiom need not hold, making it an attractive choice for modelling rankings which may contain context-dependent relations.

In Seshadri et al. [2020], maximum likelihood estimates(MLEs) are used to estimate the parameters of the CRS model. This report looks to perform the first (equal) Bayesian inference on the CRS

model, and to compare this with Bayesian inference for the Plackett-Luce model on both simulated data, and race finishing positions from the 2021 Formula 1 season.

2 Our contribution

In this paper, we provide the first (equal) Bayesian analysis of the CRS model for rankings. We use MH MCMC to sample from the posterior distribution. We perform inference on synthetic data to ensure correctness of our implementation and success of true model parameter recovery. We also analyse Formula 1 race finishing positions from 2021, and use this to allow a comparison of CRS to the VSP model [Jiang et al. [2023]].

3 Ranking data

Here we explain some concepts regarding ranking data, as well as defining the notation used throughout this paper.

Let $A = \{a_1, ..., a_N\}$ be a set of N actors and let $O = \{S \subseteq A : |S| \geq 2\}$. We are interested in viewing ordered lists of subsets of the actors in A in which the position in the list denotes a preference. We refer to these ordered lists as *rankings*.

We can formally define a ranking of the elements of a set $S \in O$ with $|S| = n$ as a function $\sigma : S \to \{1, 2, ..., n\}$ such that $\sigma(a_i) = m$ means that actor a_i is in position m in the ranking. A ranking $y = (y_1, y_2, ..., y_n)$ then solves $\sigma(y_1) = 1, ..., \sigma(y_n) = n$.

Given $i \in S$, $P(i|S)$ is the probability that i is chosen from S.

In both the Plackett-Luce and CRS models, the generation of a ranking can be viewed as a process of repeated selection from the top of the ranking down: first an actor a_i is chosen from S to be in the first position, then an actor a_j is chosen from $S \setminus a_i$ to be in the second position, and so on until the ranking is complete. Each of these individual choices are thought to be independent (the probability that i is chosen from $S \setminus j$ is independent from the probability that j is chosen from S), hence the probability of a given ranking is then the product of the probabilities of each choice:

$$P(Y|\theta) = P(y_1|\theta, \{y_1, ..., y_n\}) \times P(y_2|\theta, \{y_2, ..., y_n\}) \times ... \times P(y_{n-1}|\theta, \{y_{n-1}, y_n\}) \quad (1)$$

We assume that there is some unknown "true" ordering between the actors, and this is what we wish to infer.

3.1 Luce's choice axiom

Luce's choice axiom [Luce [2012]] is an important concept in the study of ranking data. Informally, Luce's choice axiom holds if the probability of selecting actor i over actor j from a set S is not affected by the presence or absence of any other elements in S. Formally:

Axiom 1. *A choice system on A satisfies Luce's choice axiom if for any $i, j \in A$ and $S_1, S_2 \subseteq A$ with $i, j \in S_1, S_2$:*

$$\frac{P(i|S_1)}{P(j|S_1)} = \frac{P(i|S_2)}{P(j|S_2)} \quad (2)$$

This is sometimes referred to as *independence from irrelevant alternatives (IIA)* and implies context independence of choices. Luce's choice axiom implies transitivity of preferences: if A is preferred to B, and B is preferred to C, then A will be preferred to C. The satisfaction of this axiom is a necessary condition for the use of the MNL and therefore Plackett-Luce, however intransitivity of choices has been found in areas such as sport [Chen and Joachims [2016]] and consumer preferences [Guadalupe-Lanas et al. [2020]], and hence models that do not require the restrictions of Luce's choice axiom are desired.

4 Comparison of models

4.1 A context independent model

4.1.1 Multinomial Logit Model

The Multinomial Logit Model (MNL) is used to model choosing an actor from a set of possible choices. A utility λ_i is assigned to each $i \in A$, such that a higher utility corresponds with that actor being more preferable. Let S $\subseteq$ A such that $i \in A$. Then the probability of choosing i from S is:

$$P\left(i|S\right) = \frac{\exp(\lambda_i)}{\sum_{k \in S} \exp(\lambda_k)} \tag{3}$$

As previously stated, the MNL requires satisfaction of Luce's choice axiom and is therefore context independent.

4.1.2 Plackett-Luce model

The Plackett-Luce (PL) model [Plackett [1975]] for rankings views a ranking as a top down choice problem, and models each choice using the MNL. The PL model is therefore also context independent. Let $y = (y_1, y_2, ..., y_n)$ be a ranking of the elements in the set S, such that $\sigma(y_i) = i$. Let λ = $\{\lambda_{y_1}, \lambda_{y_2}, ..., \lambda_{y_n}\}$ be the associated parameters for each actor under the Plackett-Luce model. Combining (1)and (3), the probability of y given λ is then:

$$P_{PL}\left(y|\,\lambda\right) = \prod_{i=1}^{n-1} \frac{\exp(\lambda_{y_i})}{\sum_{k=i}^{n} \exp(\lambda_{y_k})} \tag{4}$$

4.2 A context dependent model

4.2.1 CDM model

The context-dependent utility model (CDM) [Seshadri et al. [2019]] is a generalisation of the MNL that allows an escape from the confines of Luce's choice axiom by allowing the utility of an actor to change dependent on which other actors are present. In its full form, for a set of N actors this model is parameterised by an N x N matrix U, where the diagonal entries are undefined and so can be set to 0 so that U can still be viewed as an N x N matrix. There is also a low-rank factorisation of the CDM, as mentioned in Seshadri et al. [2019], however we do not examine it in this report. Under the CDM, the probability of choosing i from S is:

$$P(i|S) = \frac{\exp(\sum_{j \in S \setminus \{i\}} u_{ij})}{\sum_{k \in S} \exp(\sum_{j \in S \setminus \{k\}} u_{kj})} \tag{5}$$

4.2.2 Contextual Repeated Selection(CRS) Model

The contextual repeated selection (CRS) model [Seshadri et al. [2020]] for rankings views a ranking as a top down choice problem, and models each choice using the CDM, meaning this model is also context independent. Let U be the associated N x N CRS parameter matrix for the actors in A. Combining (1) and (5), the probability of Y given U is:

$$P_{CRS}\left(y|\,U\right) = \prod_{i=1}^{n} \frac{\exp(\sum_{k=i+1}^{n} u_{y_i, y_k})}{\sum_{j=i}^{n} \exp(\sum_{k \in \{i,...,n\} \setminus \{j\}} (u_{y_j, y_k})} \tag{6}$$

4.2.3 Plackett-Luce represented by CRS

In Seshadri et al. [2020], it is stated that the CRS model subsumes the Plackett-Luce model, however it is never explicitly stated how exactly a PL model can be represented by a CRS model. The replication code for the simulations performed in Seshadri et al. [2020] is publicly available on Github, and

through reading this I saw that to represent a PL model with parameters $\lambda = \{\lambda_1, ..., \lambda_N\}$ in a CRS model they set the CRS parameter matrix as:

$$U = \begin{pmatrix} 0 & -\lambda_2 & \cdots & -\lambda_N \\ -\lambda_1 & 0 & \cdots & -\lambda_N \\ \vdots & \vdots & \vdots & \vdots \\ -\lambda_1 & -\lambda_2 & \cdots & 0 \end{pmatrix} \tag{7}$$

We prove that this representation in the CRS model is equivalent to the PL model in the supplementary material.

This suggests that if the matrix U were to be estimated through MCMC, then λ_i can be estimated by calculating:

$$\hat{\lambda}_i = -\frac{1}{N-1} \sum_{j \in C \setminus i} u_{ji} \tag{8}$$

5 Discussion of CRS model

5.1 What does it mean for i to "beat" j?

When thinking of the underlying ordering that we expect the elements in A to have, we naturally come across the question of what it means for $i \in A$ to "beat" $j \in A$.

In Plackett-Luce, this notion seems trivial: i beats j if $\lambda_i > \lambda_j$. In this scenario, i will always be preferred over j when a choice is made from any set containing them both. This then makes it easy to form a graphical representation of the underlying structure between actors, as we can represent them as a total order using their ordered score values.

In CRS however this notion becomes more complex. The introduction of context dependence means that whether i is preferred over j depends on the set in which the preference is being made. We can therefore have multiple notions of what it means for i to "beat" j.

In the "local" context, we could say that i "beats" j if $P(i|\{i, j\}) \geq 0.5$, i.e. if i is more likely to be chosen from the set containing only i and j.

In the "global" context, we could say that i "beats" j if the probability that i is ranked higher than j in a given list is ≥ 0.5. To define this better, imagine we are only dealing with full length lists. We could say $P(i \text{ beats } j) = P(\sigma(i) < \sigma(j)) = \sum P(i|S)P(S)$ where the sum is over $S \subseteq A$, such that $i, j \in S$. So the probability that i beats j in the global context is the sum of the probability that i beats j in all settings where i and j are present multiplied by the probability of each setting occurring.

The probability of each setting occurring, P(S) would be the probability the elements in S are together in a list of that size, i.e. that the elements in $S \setminus A$ have already been selected before them.

These global probabilities for the CRS model are very computationally challenging, and so it seems more appropriate to estimate them through simulation. To do this, we simulate a large number of lists under the CRS model and store the number of lists in which i beats j. The number of lists in which i beats j divided by the number of lists simulated will be the estimated probability that i "beats" j in the global sense.

5.2 Circumstances needed for the recovery of intransitive relations

One of the main selling points of CRS is its ability to model intransitivities between actors. For example, imagine a set of 3 actors $A = \{a, b, c\}$ with a true CRS parameter matrix:

$$\begin{pmatrix} 0 & 41.75 & -58.25 \\ -8.75 & 0 & 41.75 \\ -7.75 & -8.75 & 0 \end{pmatrix} \tag{9}$$

The "local" probability that i beats j, as defined in Section 5.1, for each $i, j \in A, i \neq j$, are displayed in the matrix below to 20 decimal places of accuracy:

$$\begin{pmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix} \tag{10}$$

The result of this is an intransitive relation between all three actors in the local context: b is selected over a in the set {a,b}, a is selected over c in the set {a, c}, but c is selected over b in the set {c, b}.

Now we simulate 1 million lists using the CRS model and the true parameters to estimate the "global" probabilities that i beats j. In every one of these lists, b is ranked first, a is ranked second and c is ranked third. We therefore have no information from this dataset of full lists that c would be selected over b in the local setting, i.e. no evidence that $P(c|\{c, b\}) > 0$. This motivates the idea that to infer context dependence in the CRS model, i.e. a variation in the relative probability of selecting an actor depending on the set, we may need to view the actor in different contexts.

6 Bayesian Inference for PL and CRS

We wish to give the first (equal) Bayesian inference of the CRS model. We also give Bayesian inference of the PL model for purposes of comparison. In our experiments, we sample from the posterior of each model using Metropolis-Hastings (MH) MCMC. The MCMC algorithms for each model are given in the supplementary material.

Let $Y = \{Y_1, ..., Y_m\}$ be a set of m independent rankings, $Y_i = (y_{i,1}, ..., y_{i,|b_i|})$, and $B = \{b_1, ..., b_n\}$ such that b_i is the set of actors present in Y_i and $|b_i| \geq 2 \ \forall \ i \in \{1, ..., m\}$.

6.1 PL prior and posterior

Since the rankings in Y are assumed to be independent, the likelihood of Y given λ under the Plackett-Luce model is:

$$P(Y|\lambda) = \prod_{i=1}^{m} P_{PL}(Y_i|\lambda) \tag{11}$$

where $P_{PL}(Y_i|\lambda)$ is defined as in Section 4.1.2.

In our inference, we use a prior of $\lambda_i \sim N(0, \sigma^2_{PL})$. Let $\pi(\lambda_i)$ denote the prior of λ_i. We assume prior independence of the λ_i's and hence the prior for λ is $\pi(\lambda) = \Pi_{i=1}^{N} \pi(\lambda_i)$.

The posterior for the PL model is therefore:

$$\pi(\lambda|Y) \propto \left(\prod_{i=1}^{N} \pi(\lambda_i) \right) \times P(Y|\lambda) \tag{12}$$

6.2 CRS prior and posterior

Since the rankings in Y are assumed to be independent, the likelihood of Y given U under the CRS model is:

$$P(Y|U) = \prod_{i=1}^{m} P_{CRS}(Y_i|U) \tag{13}$$

where $P_{CRS}(Y_i|U)$ is defined as in Section 4.2.2.

In our inference, we use a prior of $u_{ij} \sim N(0, \sigma^2_{CRS})$, when $i \neq j$. If CRS is being compared to the PL model, we choose $\sigma^2_{CRS} = \frac{1}{N-1}\sigma^2_{PL}$, so that the variance of a row sum in the CRS matrix is equivalent to that of a PL parameter. This is desired so that the odds estimates are approximately the same in PL and CRS. We assume prior independence of the u_{ij}'s.

The posterior for the CRS model is therefore:

$$\pi(U|Y) \propto \left(\prod_{j=1, j\neq i}^{N} \prod_{i=1}^{N} \pi(u_{ij}) \right) \times P(Y|U) \tag{14}$$

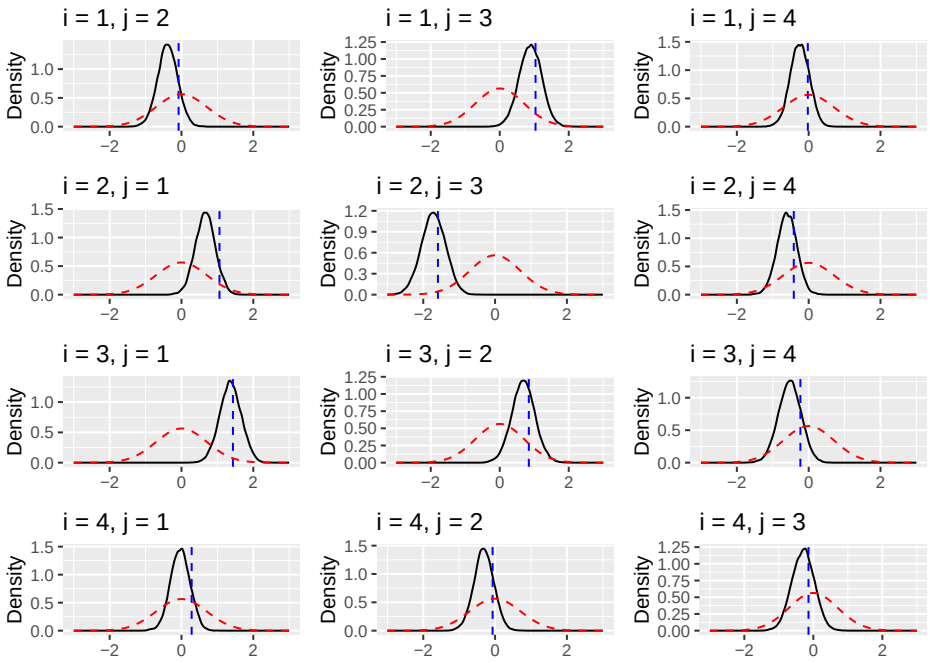

Figure 1: Density plots of $\pi(u_{ij}|X)$ in black with true u_{ij} value shown in blue and prior density shown in red

6.3 Test of CRS and PL MCMC on CRS synthetic data

To test our Bayesian inference, we generated a large amount of synthetic data (1000 rankings) from a CRS model of 4 actors and performed our PL and CRS MCMC. This is done so that we can ensure that for a large number of lists, our CRS MCMC gives us good recovery of the true parameter values. We perform MCMC under the PL model so that we can ensure our model comparison correctly selects the CRS model as preferable when the rankings are generated from it.

To generate the CRS data, a parameter matrix U was first generated such that $u_{ij} \sim N(0, 0.5)$ for $i \neq j$ and $u_{ii} = 0$. 1000 rankings of length 4 were then generated under the CRS model with parameter matrix U. Let $Z = \{Z_1, ..., Z_{1000}\}$ denote this set of rankings.

MH MCMC was performed with priors as given in Section 6 and prior variances of $\sigma^2_{CRS} = 0.5$ and $\sigma^2_{PL} = 1.5$. Each run was M = 200,000 steps with every 5th step recorded.

The MCMC for PL resulted in a minimum ESS of 21, suggesting poor convergence, whilst the MCMC for CRS resulted in a minimum ESS of 415, suggesting good convergence. The densities of the posteriors of the u_{ij}'s are given in Figure 1. These appear to show that the posterior correctly converges to the true CRS parameter values.

For model comparison, we use the WAIC [Watanabe [2013]]. Table 1 gives the $elpd_{WAIC}$ values for each model fitted to the synthetic dataset. The $elpd_{WAIC}$ is higher for the CRS model, suggesting this model is the best fit to the data, as we would expect, given the data is generated under this model.

Table 1: $elpd_{WAIC}$ values for each model on the synthetic dataset, with standard error in brackets

Plackett-Luce	CRS
-2656.1 (27.8)	**-2472.6 (27.4)**

We can conclude from this test that our CRS MCMC correctly converges to the true parameter values for long MCMC runs and large datasets, and that our model selection process correctly selects CRS over PL when using synthetic CRS data.

7 Analysis of Formula 1 data

The Formula 1 data consists of the finishing positions of drivers in each of the 22 Grand Prix's held in the 2021 season. Let $Y = \{Y_1, ..., Y_{22}\}$ denote this set of 22 rankings. 21 drivers took part in this season, however not all drivers participated in all races, and some drivers did not finish a given race. Drivers who did not finish a race were removed from the ranking for that race. Rankings are therefore of unequal lengths.

Each driver is given a unique Driver ID, an integer between 1 and 21, and it is the rankings of the Driver ID's that are used in the analysis. Let $F = \{1, 2, ..., 21\}$ be the set of Driver ID's, and hence the set of actors in the context of our models.

This dataset is the same as that analysed in Jiang et al. [2023] in which the authors fitted multiple models to the data, including Plackett-Luce and VSP, but not CRS.

We wish to understand the underlying hierarchy amongst drivers.

7.1 Plackett-Luce Model fitting

Let $\lambda = \{\lambda_1, \lambda_2, ..., \lambda_{21}\}$ be the score values for the Plackett-Luce model fitted to the Formula 1 data, Y, such that λ_i is the score value for actor $i \in F$.

We perform Bayesian inference as set out in section 6.1, with $\sigma^2_{PL} = 2$. We use MH MCMC to sample from the posterior distribution $\pi(\lambda|Y)$.

The MCMC is run for 50,000 steps, with every 5th step recorded, resulting in an output of M = 10,000 samples from the posterior $\pi(\lambda|Y)$. The first K = 1,000 samples were discarded to account for burn-in. Figure 2 shows posterior densities $\pi(\lambda_i|Y)$ for $i \in \{1, ...9\}$.

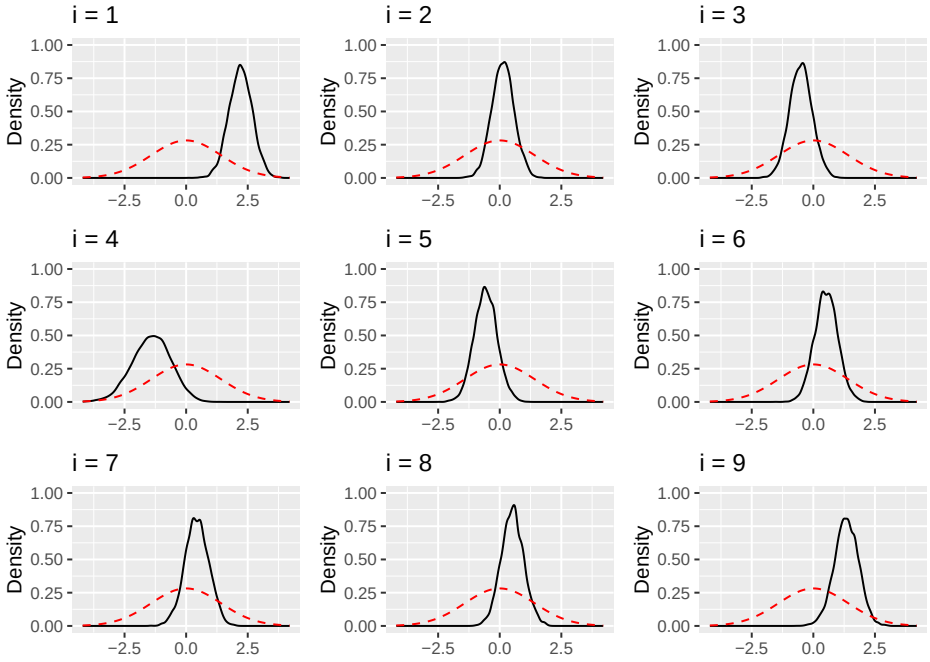

Figure 2: Density plots of $\pi(\lambda_i|X)$ in black with prior density shown in red

The minimum Effective Sample Size (ESS) of $\lambda_i \in \lambda$ was 430, suggesting good convergence of the MCMC and reasonable precision for estimates of the posterior mean.

The true parameter values are estimated by taking the mean of the posterior samples:

$$\hat{\lambda}_i = \sum_{k=B+1}^{M} X_i^{(k)} \tag{15}$$

where $X^{(k)}$ is the kth sample in the MCMC chain.

These estimated parameter values can then be used to estimate probabilities under the PL model. These probabilities can then be used to form an estimated order relation between drivers.

Due to the assumption of Luce's choice axiom, we know $P(\sigma(i) < \sigma(j) = P(i|\{i,j\})$, hence we calculate $P(i|\{i,j\})$ for each $i,j \in A$ and create a directed graph using a posterior probability threshold of 0.75. The transitive reduction of the resultant graph is shown in Figure 3(a).

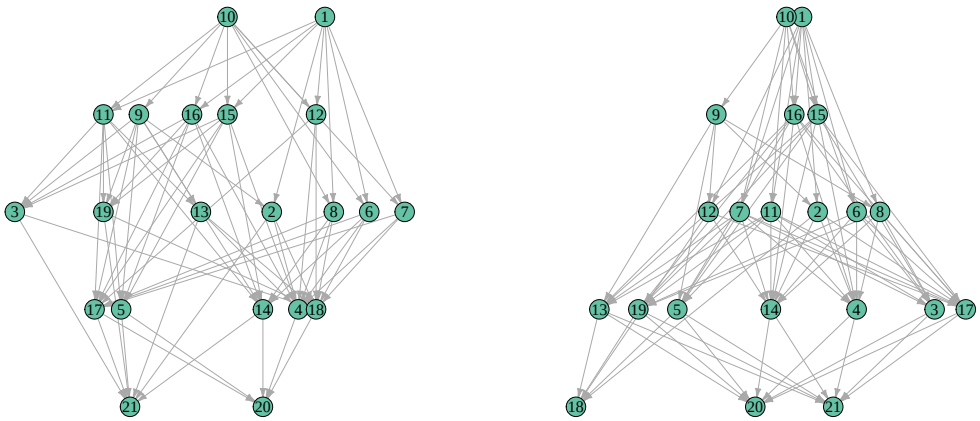

Figure 3: Directed Graphs displaying estimated order relations with posterior support > 0.75. Transitivity may be assumed. (a) Plackett-Luce, (b) CRS

7.2 CRS Model fitting

Let $U \in \mathbb{R}^{21}$ be the parameter matrix for the CRS model fitted to the Formula 1 data, Y.

We perform Bayesian inference as set out in section 6.2, with $\sigma^2_{CRS} = 0.1$. We use MH MCMC to sample from the posterior distribution $\pi(U|Y)$.

The MCMC is run for 50,000 steps, with every 5th step recorded, resulting in an output of M = 10,000 samples from the posterior $\pi(U|Y)$. The first K = 1,000 samples were discarded to account for burn-in. Figure 4 shows posterior densities for $\pi(u_{ij}|Y)$, $i,j \in \{1,...,4\}, i \neq j$.

The minimum Effective Sample Size (ESS) of $u_{ij} \in U$ was 4032, suggesting very good convergence of the MCMC.

The true parameter values are estimated by taking the mean of the posterior samples:

$$\hat{u}_{ij} = \sum_{k=B+1}^{M} X_{ij}^{(k)} \tag{16}$$

where $X^{(k)}$ is the kth recorded sample in the MCMC chain.

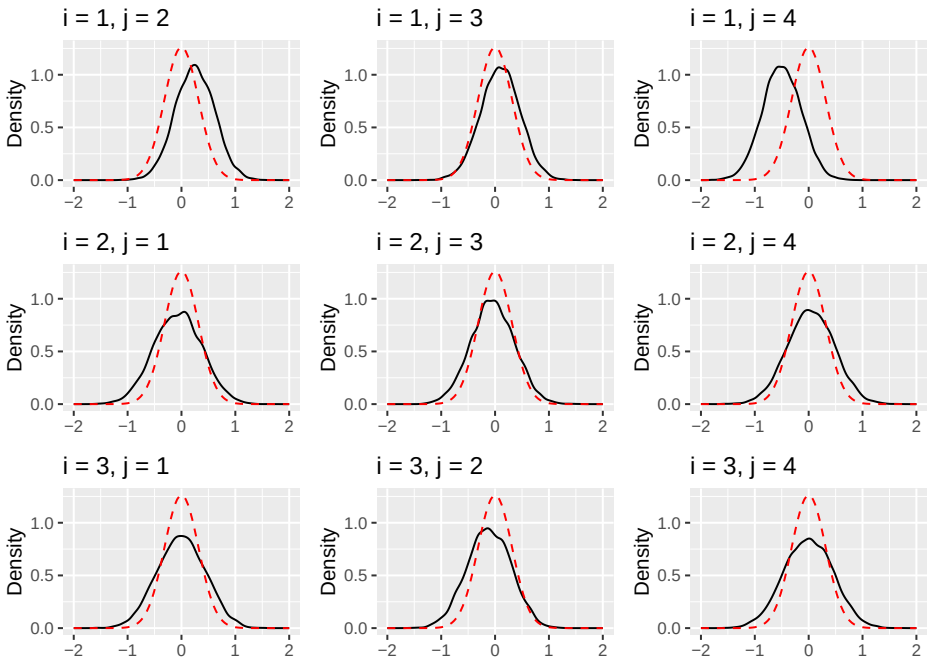

Figure 4: Density plots of $\pi(u_{ij}|X)$ in black with prior density shown in red

These estimated parameter values can then be used to estimate probabilities under the CRS model. We estimate the global probability that i beats j, $P(\sigma(i) > \sigma(j))$, for all $i \neq j$ using $\hat{U}$. These probabilities can then be used to form an estimated order relation between drivers. Figure 3(b) shows the directed graph formed using a probability threshold of 0.75.

7.3 Model comparison

We perform model comparison for the Formula 1 data using the WAIC [Watanabe [2013]], the values of which are shown in Table 1. We see that CRS performs worst out of all three models, with VSP being the most preferred.

Table 2: $elpd_{WAIC}$ values for each model on the F1 dataset, with standard error in brackets

Plackett-Luce	CRS	VSP
-652.7 (31.5)	-721.7 (25.2)	**-597.1 (25.2)**

8 Conclusion

We provide the first (equal) Bayesian inference for the CRS model for ranking data. We sampled from the posterior using MH MCMC, and performed multiple tests on this to ensure its correctness. Rejection sampling (see supplementary material) confirmed the correctness of our MCMC algorithm for both PL and CRS models, and a test of CRS MCMC on a large synthetic dataset showed good recovery of true parameter values of CRS models. In a comparison with PL MCMC on synthetic CRS data, we were able to show that our model selection using WAIC correctly selects CRS as being the preferred model when it is the true generative model of the data.

We provide an analysis of Formula 1 ranking data, in which we compared PL, CRS and VSP and found VSP to be the preferred model, and CRS to be the least preferred model.

In this report, we only worked with the full form of the CDM and CRS models. In Seshadri et al. [2020], it is seen that the low-rank factorisation of the CRS model vastly outperforms its full form, so a natural next step would be to repeat this Bayesian inference using the low-rank factorisation.

Another future topic of interest could be the underlying structure that the CRS model represents, and how this can be interpreted in an intuitive and practical manner. As discussed in section 5.1, there are multiple ways in which we can think of ordering actors under the CRS model, and it would be interesting to look at this in further detail and come up with better, fuller, but still computationally efficient, representations.

The interpretation of the CRS parameters would also be an interesting topic of further research. In Seshadri et al. [2019], the parameters of the CDM matrix are defined in terms of innate and contextual utility components, and it would be very interesting to investigate the potential recovery of these components. It would also be useful to see if the recovery of these componenets can give a measure of the impact context dependence has on a model.

References

Anna E Bargagliotti, Susan E Martonosi, Michael E Orrison, Austin H Johnson, and Sarah A Fefer. Using ranked survey data in education research: Methods and applications. *Journal of School Psychology*, 85:17–36, 2021.

Shuo Chen and Thorsten Joachims. Modeling intransitivity in matchup and comparison data. In *Proceedings of the ninth acm international conference on web search and data mining*, pages 227–236, 2016.

Jorge Guadalupe-Lanas, Jorge Cruz-Cárdenas, Verónica Artola-Jarrín, and Andrés Palacio-Fierro. Empirical evidence for intransitivity in consumer preferences. *Heliyon*, 6(3), 2020.

Chuxuan Jiang, Geoff K. Nicholls, and Jeong Eun Lee. Bayesian inference for vertex-series-parallel partial orders. In *The 39th Conference on Uncertainty in Artificial Intelligence*, 2023. URL `https://openreview.net/forum?id=K5WX50tp2ZV`.

R Duncan Luce. On the possible psychophysical laws. *Psychological review*, 66(2):81, 1959.

R Duncan Luce. *Individual choice behavior: A theoretical analysis*. Courier Corporation, 2012.

David M Pennock, Eric Horvitz, C Lee Giles, et al. Social choice theory and recommender systems: Analysis of the axiomatic foundations of collaborative filtering. *AAAI/IAAI*, 30:729–734, 2000.

Robin L Plackett. The analysis of permutations. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 24(2):193–202, 1975.

Arjun Seshadri, Alex Peysakhovich, and Johan Ugander. Discovering context effects from raw choice data. In *International Conference on Machine Learning*, pages 5660–5669. PMLR, 2019.

Arjun Seshadri, Stephen Ragain, and Johan Ugander. Learning rich rankings. In H. Larochelle, M. Ranzato, R. Hadsell, M.F. Balcan, and H. Lin, editors, *Advances in Neural Information Processing Systems*, volume 33, pages 9435–9446. Curran Associates, Inc., 2020. URL `https://proceedings.neurips.cc/paper/2020/file/6affee954d76859baa2800e1c49e2c5d-Paper.pdf`.

Sumio Watanabe. A widely applicable bayesian information criterion. *Journal of Machine Learning Research*, 14(1):867–897, 8 2013. ISSN 15324435. URL `http://arxiv.org/abs/1208.6338`.

Tongwu Zhang, Philippe Joubert, Naser Ansari-Pour, Wei Zhao, Phuc H Hoang, Rachel Lokanga, Aaron L Moye, Jennifer Rosenbaum, Abel Gonzalez-Perez, Francisco Martinez-Jimenez, et al. Genomic and evolutionary classification of lung cancer in never smokers. *Nature genetics*, 53(9): 1348–1359, 2021.