Psychometrika

Bayesian Plackett-Luce mixture models for partially ranked data

--Manuscript Draft--

Manuscript Number:	PMET-D-15-00046R2
Full Title:	Bayesian Plackett-Luce mixture models for partially ranked data
Abstract:	The elicitation of an ordinal judgment on multiple alternatives is often required in many psychological and behavioral experiments to investigate preference/choice orientation of a specific population. The Plackett-Luce model is one of the most popular and frequently applied parametric distributions to analyze rankings of a finite set of items. The present work introduces a Bayesian finite mixture of Plackett-Luce models to account for unobserved sample heterogeneity of partially ranked data. We describe an efficient way to incorporate the latent group structure in the data augmentation approach and the derivation of existing maximum likelihood procedures as special instances of the proposed Bayesian method. Inference can be conducted with the combination of the Expectation-Maximization algorithm for maximum \textit{a posteriori} estimation and the Gibbs sampling iterative procedure. We additionally investigate several Bayesian criteria for selecting the optimal mixture configuration and describe diagnostic tools for assessing the fitness of ranking distributions conditionally and unconditionally on the number of ranked items. The utility of the novel Bayesian parametric Plackett-Luce mixture for characterizing sample heterogeneity is illustrated with several applications to simulated and real preference ranked data. We compare our method with the frequentist approach and a Bayesian nonparametric mixture model both assuming the Plackett-Luce model as mixture component. Our analysis on real data sets reveals the importance of an accurate diagnostic check for an appropriate in-depth understanding of the heterogenous nature of the partial ranking data.
Keywords:	ranking data; Plackett-Luce model; Mixture Models; data augmentation; MAP estimation; Gibbs sampling; label-switching; goodness-of-fit
Article Type:	Original Research
Section/Category:	Theory and Methods (T&M)
Corresponding Author:	Cristina Mollica, Ph.D. Sapienza Università di Roma Rome, Rome ITALY
Corresponding Author's Institution:	Sapienza Università di Roma
Corresponding Author E-Mail:	cristina.mollica@uniroma1.it
Order of Authors:	Cristina Mollica, Ph.D.
	Luca Tardella, Ph.D.
Author Comments:	Dear Editor, please find enclosed our revised manuscript and a letter in which we have replied specifically to each point raised by all the reviewers and the AE. We have explicitly pointed out where new material and further comments have been added. Thanks a lot for the opportunity to submit our revision, for your comments and suggestions. Thanks also for all the remarks provided by the AE and the reviewers. We hope you will find our revised paper suitable for publication. Sincerely yours, Cristina Mollica and Luca Tardella

Response to Reviewers:	Please, find enclosed PDF letters for Editor, AE and reviewers.
Additional Information:	
Question	**Response**
If your article is accepted for publication or a reviewer requests electronic supplementary material, are you willing to share any related data and/or code to accompany your article? In the interest of transparency and reproducibility of results, Psychometika strongly encourages authors to include with published papers, where possible, electronic supplementary materials which will be published online on SpringerLink with the article.	Yes
Please specify the type of supplementary material you will be able to provide. Data and code should be clearly documented to allow replication and verification of the results presented in the final version of the paper. as follow-up to "If your article is accepted for publication or a reviewer requests electronic supplementary material, are you willing to share any related data and/or code to accompany your article? In the interest of transparency and reproducibility of results, Psychometika strongly encourages authors to include with published papers, where possible, electronic supplementary materials which will be published online on SpringerLink with the article. "	Supplementary Tables and Figures

Reply to Editor

Dear Editor,

please find enclosed our revised manuscript and a letter in which we have replied specifically to each point raised by all the reviewers and the AE. We have explicitly pointed out where new material and further comments have been added.

Thanks a lot for the opportunity to submit our revision, for your comments and suggestions. Thanks also for all the remarks provided by the AE and the reviewers. We hope you will find our revised paper suitable for publication.

Sincerely yours,

Point-to-point reply to AE and referees

Associate Editor

We would like to thank all the referees and the Associate Editor for their helpful comments and suggestions which greatly improved our originally submitted manuscript.

Before replying to all the critical points, we list below the main changes of this new revised version of the paper:

a) the former simulation study has been extended to account for cases where most of the observed orderings are strictly partial. We have discussed the effect of the censoring mechanism on the performance of all the competing methods and included the evidence on the fitting diagnostics (subsection 4.1);

b) with the results of the new simulation we have provided further evidence of the better performance of DIC_1 and PBIC_1 over the other competing criteria in the fourth mixture scenario with $G^* = 4$ (subsection 4.1) and we have provided some discussion on the comparison with BIC (section 5);

c) in order to make notation more rigorous, we have modified the normalization term appearing in the denominator of the PL distribution (section 2 and 3). Since the actual length of a generic partial ordering $\pi_s^{-1} = (\pi_s^{-1}(1), \dots, \pi_s^{-1}(n_s))$ is $n_s \leq K$, we have replaced the former expressions of the normalization term, given by

$$\sum_{\nu=t}^{K} p_{\pi_s^{-1}(\nu)} \qquad \text{and} \qquad \sum_{\nu=t}^{K} p_{g\pi_s^{-1}(\nu)}$$

(respectively for the homogeneous PL and the PL mixture). We have preferred the notation

$$\sum_{i=1}^{K} p_i - \sum_{\nu=1}^{t-1} p_{\pi_s^{-1}(\nu)} \qquad \text{and} \qquad \sum_{i=1}^{K} p_{gi} - \sum_{\nu=1}^{t-1} p_{g\pi_s^{-1}(\nu)}.$$

For the mixture case, we have modified accordingly the definition of the binary array δ_{sti} (subsection 3.2), as described in the manuscript.

Comments

1. *All the referees found the revised version was substantially improved. The authors have addressed most of the previous concerns. However, there are still several issues that need to be taken into account. I endorse referee 1's comment about the simulation study. It is of interest to consider the case where most of the observations are partially ranked.*

   Please, see reply to point #1 of Reviewer #1.

2. *Also, you did not explain why DIC_1 and $BPIC_1$ performed better than others when the number of PL components is four. A discussion or explanation on this finding is welcome.*

   Please, see reply to point #5 of Reviewer #1.

3. *Last but not least, many expressions are hard to understand. For instance, on page 4, the first paragraph: "Hence, the ability of the BNPPLM of identifying a suitable finite number of clusters underlying…" Such unclear descriptions can be found here and there; see also referee 1's report. I suggest the authors carefully edit the manuscript to improve its readability and succinctness.*

   On page 4, we have rephrased the original sentence and split it into shorter multiple sentences as follows:

   " Hence, the ability of the BNPPLM of identifying a suitable finite number of clusters underlying the observed data is related to the random partition associated to the sequential draw of partial rankings. In fact, for each sample unit the partial ranking is generated from the corresponding random vector of support parameters, which in turn follows a Dirichlet allocation model (McCullagh et al., 2008). Multiple sample units can then share the same parameter vector and, hence, belong to the same group of the partition. One can rely on the posterior simulation of the parameters and use, as suggested by Caron et al. (2014), the ad-hoc method originally proposed by Dahl (2006) to estimate a suitable finite number of underlying groups."

   To make the paper more readable, we have also shortened and rephrased most of the longest sentences. Finally, some typos have been also removed.

Reviewer #1

We would like to thank all the referees and the Associate Editor for their helpful comments and suggestions which greatly improved our originally submitted manuscript.

Comments

1. *This is review for manuscript PMET-D-15-00046R1. First, I would like to thank the authors for their responses to my previous comments. While I am satisfied with most of the responses, I also have some other comments and suggestions for this revision. My main concern is about the simulation study. While the authors have made great improvement in this section, I have more suggestions on how to make the simulation study more comprehensive and useful. First, in these simulation studies, most of the data are fully ranked. It would be very useful to consider situations where most of the data are partially ranked.*

The former simulation study has been extended to account for cases where most of the orderings are strictly partial. In the updated subsection 4.1, we have considered two additional censoring settings for the random truncation of the simulated complete orderings, both leading to samples with more than 50% of strictly partial observations. We have then discussed the effect of the censoring mechanism on the performance of all the competing methods with the relative merits.

2. *Second, it would also be very helpful if the authors can incorporate some analysis in the simulation studies about model assessment presented in section 3.5.*

   We have incorporated in the simulation study (subsection 4.1) some evidence on the diagnostic tools proposed in subsection 3.5 to evaluate the effectiveness of the posterior predictive check. For the sake of brevity, only the results for the unconditional fitting measures ($p_B(1)$ and $p_B(2)$) in most difficult-to-estimate mixture population scenario with $G^* = 4$ components (Scenario 4) are reported. The results for the other population scenarios and the conditional fitting diagnostics ($\tilde{p}_B(1)$ and $\tilde{p}_B(2)$) provide similar evidence and are not reported.

3. *On page 16, line 13, it said "In scenario 1, BPIC$_1$, BPIC$_2$, BICM$_1$ and BICM$_2$ always recover the actual absence of a group structure." It is not clear what is "the actual absence of a group structure". Does it mean the actual number of latent group? If so, BICM$_2$ does not recover all the actual number (99 out of 100).*

   The expression "the actual absence of a group structure" means the actual absence of heterogeneity which corresponds to $G^* = 1$. On page 16 lines 9-11, we have rephrased this statement as follows:

   "Regarding censoring setting A, in Scenario 1 BPIC$_1$, BPIC$_2$ and BICM$_1$ always recover the actual absence of heterogeneity (i.e. $G^* = 1$).'

4. *On page 16, line 52, it said "The vector (42, 17, 0, 29, 62) lists the number of missing responses for each item ..." Should the vector includes 6 elements since there are 6 items?*

   Yes, it should. Thanks for pointing this out. In fact, we missed to report the last entry (= 27). We have now fixed this by adding it to the vector (subsection 4.2, on page 18, line 12).

5. *It would be very helpful if the authors can provide some discussions about why BPIC$_1$ and DIC$_1$ performs better in the simulation study. It is also worth discuss why BIC performs very well except the last scenario.*

   We have provided the following discussion in the concluding section:

"Additionally, this work provided some incremental findings on the performance of many alternative Bayesian selection criteria. Our investigation suggests, besides the most frequently adopted DIC_1, the use of $BPIC_1$. Also BIC performed well for smaller values of G^*. However, for larger value of G^* we confirm BIC's tendency to underestimate the true number of groups, as also pointed out in other mixture settings, see for example Celeux and Soromenho (1996); Lukočienė and Vermunt (2009) and Bulteel et al. (2013). In line with this evidence, under Scenario 4 no overestimation is present with BIC; on the other hand, BIC leads to underestimate the true number G^* of components for at least 30% of the simulated data sets in all the three censoring settings. Indeed, one could argue that, as a function of the sample size, the penalty term of BIC does not account for the varying rate of truncation, leading to a too severe penalization and, hence, to the selection of more parsimonious models. Conversely, with DIC_1 and $BPIC_1$ the effective number of parameters depends on the posterior deviance distribution, that inherently penalizes for the increasing parameterization and the higher censoring rate. For this reason, the two Bayesian criteria could be expected to return a more adaptive and suitable estimation of model complexity."

Reviewer #3

We would like to thank all the referees and the Associate Editor for their helpful comments and suggestions which greatly improved our originally submitted manuscript.

Comments

1. *I am happy with the current manuscript. I think it makes a nice contribution to the ranking literature. I would encourage the authors to make their code available for wider use.*

   Thank you for your positive comment. We plan to release an alpha version of an R package for the analysis of partially ranked data via PL mixtures and extensions thereof. Meanwhile, the undocumented R code is available upon request from the authors or can be directly downloaded from

   `https://mega.nz/#F!iJwxkQSY!Gt1VH5QYFfRP84AftmUr-Q`.

References

Bulteel, K., Wilderjans, T. F., Tuerlinckx, F., and Ceulemans, E. (2013). CHull as an alternative to AIC and BIC in the context of mixtures of factor analyzers. *Behavior Research Methods*, 45(3):782–791.

Caron, F., Teh, Y. W., and Murphy, T. B. (2014). Bayesian nonparametric Plackett-Luce models for the analysis of preferences for college degree programmes. *The Annals of Applied Statistics*, 8(2):1145–1181.

Celeux, G. and Soromenho, G. (1996). An entropy criterion for assessing the number of clusters in a mixture model. *Journal of Classification*, 13(2):195–212.

Dahl, D. B. (2006). Model-based clustering for expression data via a Dirichlet process mixture model. *Kim-Anh Do, Peter Müller & Marina Vannucci (Eds.)*, pages 201–218.

Lukočienė, O. and Vermunt, J. K. (2009). Determining the number of components in mixture models for hierarchical data. In *Advances in data analysis, data handling and business intelligence*, pages 241–249. Springer.

McCullagh, P., Yang, J., et al. (2008). How many clusters? *Bayesian Analysis*, 3(1):101–120.

BAYESIAN PLACKETT-LUCE MIXTURE MODELS FOR PARTIALLY RANKED DATA

Cristina Mollica

dipartimento di scienze statistiche, sapienza università di roma, piazzale a. moro 5, (00185) roma - italy

cristina.mollica@uniroma1.it

Luca Tardella

dipartimento di scienze statistiche, sapienza università di roma, piazzale a. moro 5, (00185) roma - italy

luca.tardella@uniroma1.it

June 1, 2016

Correspondence should be sent to Cristina Mollica

E-Mail: cristina.mollica@uniroma1.it

BAYESIAN PLACKETT-LUCE MIXTURE MODELS FOR PARTIALLY RANKED DATA

June 1, 2016

BAYESIAN PLACKETT-LUCE MIXTURE MODELS FOR PARTIALLY RANKED DATA

Abstract

The elicitation of an ordinal judgment on multiple alternatives is often required in many psychological and behavioral experiments to investigate preference/choice orientation of a specific population. The Plackett-Luce model is one of the most popular and frequently applied parametric distributions to analyze rankings of a finite set of items. The present work introduces a Bayesian finite mixture of Plackett-Luce models to account for unobserved sample heterogeneity of partially ranked data. We describe an efficient way to incorporate the latent group structure in the data augmentation approach and the derivation of existing maximum likelihood procedures as special instances of the proposed Bayesian method. Inference can be conducted with the combination of the Expectation-Maximization algorithm for maximum *a posteriori* estimation and the Gibbs sampling iterative procedure. We additionally investigate several Bayesian criteria for selecting the optimal mixture configuration and describe diagnostic tools for assessing the fitness of ranking distributions conditionally and unconditionally on the number of ranked items. The utility of the novel Bayesian parametric Plackett-Luce mixture for characterizing sample heterogeneity is illustrated with several applications to simulated and real preference ranked data. We compare our method with the frequentist approach and a Bayesian nonparametric mixture model both assuming the Plackett-Luce model as mixture component. Our analysis on real data sets reveals the importance of an accurate diagnostic check for an appropriate in-depth understanding of the heterogenous nature of the partial ranking data.

Key words: Ranking data, Plackett-Luce model, Mixture models, Data augmentation, MAP estimation, Gibbs sampling, label-switching, goodness-of-fit

1. Introduction

Choice behavior is a theme of great interest in several research areas, such as social and psychological sciences, but its investigation usually involves variables which cannot be directly observed and measured in an objective and precise manner. For this reason, the evidence in choice experiments is often collected in ordinal form, that is, in terms of *ranking data*. More specifically, ranked data arise in those studies where a sample of N people is presented a finite set of K alternatives, called *items*, and is asked to rank them according to a certain criterion, such as personal preferences or attitudes. Thus, a generic ranking is the result of a comparative judgment on the competing alternatives expressed in the form of order relation. Interest in ranked data analysis is motivated, for example, by marketing and political surveys, but also by psychological and behavioral studies consisting, for instance, in the ordering of words/topics according to the perceived association with a reference subject.

Ranked data analysis has been addressed from numerous perspectives, as revealed by a wide and consolidated literature reviewed in Marden (1995) and, more recently, in Alvo and Yu (2014). Of course, a significant role is played by the parametric modeling of ranking data, which sometimes is inspired by possible patterns underlying the (random) mechanism of formation of individual preferences. Nowadays there is a large number of parametric ranking distributions but, despite the large availability of options, often none of them is able to embody the appropriate flexibility to represent the heterogeneous nature of real data. Consequently, it is natural to extend them to the mixture context. Our work focuses on the finite mixture approach with the Plackett-Luce model (PL) as parametric component within a Bayesian inferential framework, aimed at analyzing heterogeneous partial rankings. It parallels the frequentist approach in Gormley and Murphy (2006). Recent works considering Bayesian mixture modeling based on the PL are Gormley and Murphy (2009) and Caron et al. (2014). Gormley and Murphy (2009) deals with a grade of membership model where, at each stage of the sequential ranking process, each sample unit has a specific partial membership of each component. This model is inherently different from the usual finite mixture model with discrete distributions on the latent variable developed in Gormley and Murphy (2006) and is better suited for soft clustering purposes. A

Bayesian nonparametric PL based on a Gamma process to account for an infinite number of items, shortened as BNPPL, is developed in Caron et al. (2012). The BNPPL has been subsequently extended to the mixture context, hereinafter abbreviated with BNPPLM, by Caron et al. (2014) for analyzing clustered partial ranking data. This work relies on modeling the exchangeable sequence of random partial orderings with an infinite mixture derived by means of a stick-breaking construction of the weights, corresponding to a Dirichlet process mixture. Hence, although one can consider the BNPPLM developed by Caron et al. (2014) as a natural generalization of our finite mixture framework, we point out two important differences: (i) in our parametric setting each single component is a standard PL for finite orderings (possibly truncated), whereas the BNPPL component models the orderings of a possibly arbitrary number of items; (ii) in our framework the cardinality of the mixture models is explicitly defined as finite, whereas it is infinite in Caron et al. (2014). Hence, the ability of the BNPPLM of identifying a suitable finite number of clusters underlying the observed data is related to the random partition associated to the sequential draw of partial rankings. In fact, for each sample unit the partial ranking is generated from the corresponding random vector of support parameters, which in turn follows a Dirichlet allocation model (McCullagh et al., 2008). Multiple sample units can then share the same parameter vector and, hence, belong to the same group of the partition. One can rely on the posterior simulation of the parameters and use, as suggested by Caron et al. (2014), the *ad-hoc* method originally proposed by Dahl (2006) to estimate a suitable finite number of underlying groups.

In order to address the typical issues faced with a parametric finite mixture analysis, we devote special attention to alternative criteria for the determination of the appropriate number of components. Additionally, we investigate suitable diagnostic tools to detect possible deficiencies of the PL parametric class in capturing the underlying dependence structure and highlight some critical issues in combining partial orderings characterized by a different number of ranked items. Indeed, we will show how this step is relevant for an appropriate recognition of the parsimonious group structure.

The outline of the article is the following. In Section 2 we review the PL for partial orderings and its Bayesian estimation based on data augmentation. The novel Bayesian mixture of PL and the related inferential procedures are presented in Section 3, together with alternative Bayesian model selection criteria and model assessment diagnostics. Illustrative applications of the proposed methods to both simulated and real ranking data follow in Section 4. In Section 5 the paper ends with concluding remarks and hints to future developments.

2. The Plackett-Luce model

2.1. Model specification

A ranking can be elicited through a series of sequential comparisons in which a single item is preferred to all the remaining alternatives and, after being selected, is removed from the next comparisons. This is the basic construction underlying the PL, a well-established parametric distribution among the so-called *stagewise ranking models*. It was originally introduced by Luce (1959) and Plackett (1975). More specifically, by denoting with K the total number of items to be ranked, the PL is parametrized by the *support parameters* $\underline{p} = (p_1, \ldots, p_K)$ representing positive constants associated to each item: the higher the value of the support parameter p_i, the greater the probability for the i-th item to be preferred at each selection stage. Let $\underline{\pi}^{-1} = \{\pi_s^{-1}\}_{s=1}^N$ be a random sample consisting of N partial top orderings of the form $\pi_s^{-1} = (\pi_s^{-1}(1), \ldots, \pi_s^{-1}(n_s))$. With a slight abuse of notation, n_s is the length of the s-th partial ordering, that is, the number of items ranked by unit s in the top n_s positions. The remaining $K - n_s$ items are assumed to be ranked lower. In our notation a full ordering corresponds to the case $n_s = K - 1$, since once $K - 1$ items have been ranked, the last position is automatically determined. Under the PL the contribution to the likelihood from the s-th partial ordering is given by

$$\mathbf{P}_{\mathrm{PL}}(\pi_s^{-1}|\underline{p}) = \prod_{t=1}^{n_s} \frac{p_{\pi_s^{-1}(t)}}{\sum_{i=1}^K p_i - \sum_{\nu=1}^{t-1} p_{\pi_s^{-1}(\nu)}}. \tag{1}$$

We notice that for strictly partial orderings ($n_s < K - 1$) the distribution in (1) corresponds to the marginal PL distribution for full orderings obtained by integrating out the items ranked in the last $K - n_s$ positions. An important summarizing feature of $\mathbf{P}_{\mathrm{PL}}(\cdot|\underline{p})$ is the *modal ordering* $\sigma_{\underline{p}}^{-1}$, corresponding to the ordering of the support parameters $\underline{p}$ from the largest to the smallest.

2.2. Model estimation

The main inferential issue related to formulation (1) concerns the presence of the annoying normalization term $\left(\sum_{i=1}^{K} p_i - \sum_{\nu=1}^{t-1} p_{\pi_s^{-1}(\nu)}\right)$, that does not permit the direct maximization of the likelihood. In the maximum likelihood estimation (MLE) framework, Hunter (2004) overcomes this difficulty by applying the Minorization-Maximization algorithm, an iterative optimization method relying on the replacement of the original PL log-likelihood with a minorizing surrogate objective function. In the Bayesian perspective, instead, a related efficient solution is derived by Caron and Doucet (2012), whose work can be considered the starting point of our parametric proposal presented in the next section. In particular, Caron and Doucet (2012) propose to introduce a data augmentation step with latent quantitative variables $\underline{y} = (y_{st})$ for $s = 1, \ldots, N$ and $t = 1, \ldots, n_s$, whose conditional joint distribution is given by

$$f(\underline{y}|\, \underline{\pi}^{-1}, \underline{p}) = \prod_{s=1}^{N} \prod_{t=1}^{n_s} f_{\text{Exp}}\left(y_{st} \middle| \sum_{i=1}^{K} p_i - \sum_{\nu=1}^{t-1} p_{\pi_s^{-1}(\nu)}\right), \tag{2}$$

where $f_{\text{Exp}}(\cdot|\lambda)$ denotes the Negative Exponential density with rate parameter λ. The parametric assumption (2) entails remarkable simplifications for the implementation of both the posterior optimization and the Gibbs Sampling (GS) algorithm. The success of the Bayesian device introduced by Caron and Doucet (2012) is due to the combination of (2) with a conjugate prior specification. This latter aspect moves from the *Thurstonian interpretation* of (1), that is, Thurstone's ranking model reduces to the PL when the Gumbel distribution is employed as distribution of the latent scores, see Yellott (1977). Caron and Doucet (2012) exploited the conjugacy of the Gamma density with the Gumbel distribution and derived a simple and effective GS scheme for the approximation of the posterior distribution.

3. Bayesian mixture of Plackett-Luce models

A wide variety of research contexts requires a model-based analysis accounting for the presence of differential patterns in a collection of partially ranked data. To our knowledge, Bayesian inference of a finite PL mixture has not been previously developed in the literature concerning parametric methods to analyze such data. Bayesian PL estimation appeared so far in the literature is either limited to the homogeneous case, as in Guiver and Snelson (2009) and Caron and Doucet (2012), or accounts simultaneously for an infinite mixture configuration and an infinite number of items through a nonparametric approach, see Caron et al. (2012, 2014). In the next subsections we detail the novel Bayesian PL mixture model for partial top rankings.

3.1. Model and prior specification

Let $\underline{\pi}^{-1}$ be a random sample of partial top orderings with varying lengths drawn from a G-component PL mixture, in symbols

$$\pi_1^{-1},\dots,\pi_N^{-1}|\underline{p},\underline{\omega} \overset{iid}{\sim} \sum_{g=1}^{G}\omega_g\,\mathbf{P}_{\text{PL}}(\pi_s^{-1}|\underline{p}_g),$$

where $\underline{p}_g$ is the support parameter vector specific of the g-th mixture component and ω_g is the corresponding weight. In order to suitably generalize the data augmentation approach in Caron and Doucet (2012) within the finite mixture framework, we need to introduce an additional latent feature of each generic sample unit s, represented by the unobserved group labels

$$\underline{z}_s = (z_{s1},\dots,z_{sG})|\,\underline{\omega} \overset{iid}{\sim} \text{Multinom}(1,\underline{\omega}=(\omega_1,\dots,\omega_G)),$$

whose univariate marginal distribution corresponds to a Bernoulli random variable (r.v.) such that

$$z_{sg} = \begin{cases} 1 & \text{if unit } s \text{ belongs to the } g\text{-th mixture component,} \\ 0 & \text{otherwise.} \end{cases}$$

We propose to include the unobserved group labels $\underline{z}$ in the data augmentation strategy as follows

$$f(\underline{y}|\,\underline{\pi}^{-1},\underline{z},\underline{p},\underline{\omega}) = \prod_{s=1}^{N}\prod_{t=1}^{n_s} f_{\text{Exp}}\left(y_{st}\middle|\prod_{g=1}^{G}\left(\sum_{i=1}^{K}p_{gi}-\sum_{\nu=1}^{t-1}p_{g\pi_s^{-1}(\nu)}\right)^{z_{sg}}\right). \tag{3}$$

This implies that the latent group labels determine the cluster-specific support parameters acting on the underlying quantitative variables $\underline{y}$. Once the model governing observed and latent variables is specified, a fully Bayesian approach requires the elicitation of the joint prior distribution for the unknown parameters. We choose prior distributions with independent $\underline{p}$ and $\underline{\omega}$, so that $f_0(\underline{p},\underline{\omega}) = f_0(\underline{p})f_0(\underline{\omega})$, and a convenient conjugate structure, similarly to the homogeneous population case. For the support parameters, in fact, we extend the initial distribution in Caron and Doucet (2012) by defining independent $p_{gi} \sim \mathrm{Ga}(c_{gi}, d_g)$, where the Gamma r.v.'s are indexed by the shape and the rate parameter. Finally, for the mixture weights, taking values in the $(G-1)$-dimensional simplex, we make the standard prior assumption $\underline{\omega} \sim \mathrm{Dir}(\alpha_1, \dots, \alpha_G)$.

3.2. MAP estimation

In the presence of the latent variables $\underline{y}$ and $\underline{z}$, we can construct an EM algorithm in order to optimize the posterior distribution and learn the posterior mode (MAP estimate). The complete-data likelihood can be factorized as $L_c(\underline{p},\underline{\omega},\underline{y},\underline{z}) = f(\underline{y}|\,\underline{\pi}^{-1},\underline{z},\underline{p},\underline{\omega})\,\mathbf{P}(\underline{\pi}^{-1},\underline{z}|\underline{p},\underline{\omega})$, that is, the product of the full-conditional (3) times the standard complete-data likelihood of a mixture model specification without data augmentation with $\underline{y}$. With simple algebra, both factors of the complete-data likelihood can be rearranged in order to explicit a multinomial form in $\underline{z}$ as follows

$$f(\underline{y}|\,\underline{\pi}^{-1},\underline{z},\underline{p},\underline{\omega}) = \prod_{s=1}^{N}\prod_{g=1}^{G}\left(\prod_{t=1}^{n_s}\left(\sum_{i=1}^{K} p_{gi} - \sum_{\nu=1}^{t-1} p_{g\pi_s^{-1}(\nu)}\right) e^{-\sum_{t=1}^{n_s} y_{st}\left(\sum_{i=1}^{K} p_{gi} - \sum_{\nu=1}^{t-1} p_{g\pi_s^{-1}(\nu)}\right)}\right)^{z_{sg}}$$

and

$$\mathbf{P}(\underline{\pi}^{-1},\underline{z}|\underline{p},\underline{\omega}) = \prod_{s=1}^{N}\prod_{g=1}^{G}\left(\omega_g \prod_{t=1}^{n_s} \frac{p_{g\pi_s^{-1}(t)}}{\sum_{i=1}^{K} p_{gi} - \sum_{\nu=1}^{t-1} p_{g\pi_s^{-1}(\nu)}}\right)^{z_{sg}}.$$

Hence,

$$L_c(\underline{p},\underline{\omega},\underline{y},\underline{z}) = \prod_{s=1}^{N}\prod_{g=1}^{G}\left(\omega_g \prod_{i=1}^{K} p_{gi}^{u_{si}} e^{-p_{gi}\sum_{t=1}^{n_s}\delta_{sti} y_{st}}\right)^{z_{sg}},$$

where

$$u_{si} = \begin{cases} 1 & \text{if } i \in \{\pi_s^{-1}(1), \dots, \pi_s^{-1}(n_s)\}, \\ 0 & \text{otherwise,} \end{cases}$$

and

$$\delta_{sti} = \begin{cases} 1 & \text{if } i \notin \{\pi_s^{-1}(1), \dots, \pi_s^{-1}(t-1)\}, \\ 0 & \text{otherwise,} \end{cases}$$

with $\delta_{s1i} = 1$ for all $s = 1, \dots, N$ and $i = 1, \dots, K$. We denote the complete-data log-likelihood with $l_c(\underline{p}, \underline{\omega}, \underline{y}, \underline{z}) = \log L_c(\underline{p}, \underline{\omega}, \underline{y}, \underline{z})$. The implementation of the EM algorithm in the Bayesian framework iterates: (i) the M-step, maximizing with respect to $(\underline{p}, \underline{\omega})$ the following objective function

$$Q((\underline{p}, \underline{\omega}), (\underline{p}^*, \underline{\omega}^*)) = \mathbb{E}_{\underline{y}, \underline{z} | \underline{\pi}^{-1}, \underline{p}^*, \underline{\omega}^*}[l_c(\underline{p}, \underline{\omega}, \underline{y}, \underline{z})] + \log f_0(\underline{p}, \underline{\omega});$$

(ii) the E-step, which relies on the conditional joint distribution of all the latent variables given by

$$\mathbf{P}(\underline{y}, \underline{z} | \underline{\pi}^{-1}, \underline{p}, \underline{\omega}) = f(\underline{y} | \underline{\pi}^{-1}, \underline{z}, \underline{p}, \underline{\omega}) \, \mathbf{P}(\underline{z} | \underline{\pi}^{-1}, \underline{p}, \underline{\omega}).$$

The E-step returns

$$\begin{aligned} Q((\underline{p}, \underline{\omega}), (\underline{p}^*, \underline{\omega}^*)) = &\sum_{s=1}^{N} \sum_{g=1}^{G} \hat{z}_{sg} \left(\log \omega_g + \sum_{i=1}^{K} \left(u_{si} \log p_{gi} - p_{gi} \sum_{t=1}^{n_s} \frac{\delta_{sti}}{\sum_{i=1}^{K} \delta_{sti} p_{gi}^*} \right) \right) \\ &+ \sum_{g=1}^{G} (\alpha_g - 1) \log \omega_g + \sum_{g=1}^{G} \sum_{i=1}^{K} \big((c_{gi} - 1) \log p_{gi} - d_g p_{gi} \big), \end{aligned}$$

where the posterior membership probabilities $\hat{z}_{sg}$ are obtained as

$$\hat{z}_{sg} = \frac{\omega_g^* \, \mathbf{P}_{\text{PL}}(\pi_s^{-1} | \underline{p}_g^*)}{\sum_{g'=1}^{G} \omega_{g'}^* \, \mathbf{P}_{\text{PL}}(\pi_s^{-1} | \underline{p}_{g'}^*)}.$$

Differentiating the objective function Q with respect to each p_{gi} and equating to zero yields the updated support parameters of the M-step

$$p_{gi} = \frac{c_{gi} - 1 + \hat{\gamma}_{gi}}{d_g + \sum_{s=1}^{N} \hat{z}_{sg} \sum_{t=1}^{n_s} \frac{\delta_{sti}}{\sum_{i=1}^{K} \delta_{sti} p^*_{gi}}}$$

for $g = 1, \ldots, G$ and $i = 1, \ldots, K$, where $\hat{\gamma}_{gi} = \sum_{s=1}^{N} \hat{z}_{sg} u_{si}$. Optimizing Q with respect to $\underline{\omega}$, subject to the canonical constraint $\sum_{g=1}^{G} \omega_g = 1$, yields the updated mixture weights

$$\omega_g = \frac{\alpha_g - 1 + \sum_{s=1}^{N} \hat{z}_{sg}}{\sum_{g'=1}^{G} \alpha_{g'} - G + N} \qquad g = 1, \ldots, G.$$

Notice that when $G = 1$ the MAP procedure collapses into the single updating formula obtained by Caron and Doucet (2012). Moreover, similarly to their method, also in our mixture approach we can recover the MLE as special case of the noninformative Bayesian analysis with flat priors, obtained by setting $c_{gi} = 1$, $d_g = 0$ and $\alpha_g = 1$. Such a configuration of the hyperparameters, in fact, reduces the proposed MAP estimation to the algorithm described by Gormley and Murphy (2006) in the frequentist framework.

3.3. Gibbs Sampling

In order to draw a sample from the joint posterior distribution and learn about the uncertainty associated to the final estimates, we detail the implementation of a GS procedure. The conjugate prior configuration described in Section 3.1, combined with the complete-data likelihood $L_c(\underline{p}, \underline{\omega}, \underline{y}, \underline{z})$, leads to a sampling scheme with simple parametric distributions to be drawn from. In particular, the full-conditionals of the latent component labels are easily derived by noting that $\mathbf{P}(\underline{z}|\underline{\pi}^{-1}, \underline{y}, \underline{p}, \underline{\omega}) \propto L_c(\underline{p}, \underline{\omega}, \underline{y}, \underline{z})$, implying the following multinomial structure

$$\mathbf{P}(\underline{z}_s|\pi_s^{-1}, \underline{y}_s, \underline{p}, \underline{\omega}) \propto \prod_{g=1}^{G} \left(\omega_g \prod_{i=1}^{K} p_{gi}^{u_{si}} e^{-p_{gi} \sum_{t=1}^{n_s} \delta_{sti} y_{st}} \right)^{z_{sg}}.$$

The full-conditionals of the support parameters are still members of the Gamma family with hyperparameters suitably updated as follows

$$\mathbf{P}(p_{gi}|\,\underline{\pi}^{-1},\underline{y},\underline{z},p_{[-gi]},\underline{\omega})\propto f_0(p_{gi})L_c(\underline{p},\underline{\omega},\underline{y},\underline{z})\propto p_{gi}^{c_{gi}+\gamma_{gi}-1}e^{-p_{gi}(d_g+\sum_{s=1}^{N}z_{sg}\sum_{t=1}^{n_s}\delta_{sti}y_{st})},$$

where $\gamma_{gi}=\sum_{s=1}^{N}z_{sg}u_{si}$ is the number of units belonging to cluster g who have ranked item i and $p_{[-gi]}$ denotes the matrix $\underline{p}$ of the support parameters without the (g,i)-th entry. Also the full-conditional of the mixture weights has the same form of the corresponding prior class, obtained as

$$\mathbf{P}(\underline{\omega}\,|\,\underline{\pi}^{-1},\underline{y},\underline{z},\underline{p})\propto f_0(\underline{\omega})L_c(\underline{p},\underline{\omega},\underline{y},\underline{z})\propto f_0(\underline{\omega})\,\mathbf{P}(\underline{z}|\,\underline{\omega})=\prod_{g=1}^{G}\omega_g^{\alpha_g+\sum_{s=1}^{N}z_{sg}-1}.$$

Finally, the full-conditional of $\underline{y}$ is given by construction in the assumption (3).

Note that the EM and the GS can be conveniently combined by employing the MAP solution as good initialization of the chain in the MCMC simulation. However, when one adopts an MCMC procedure to derive Bayesian approximate inference of a mixture model, the MCMC sample can be affected by the annoying identifiability issue, known as *label-switching* phenomenon (LS). This may prevent from a straightforward posterior estimation (Celeux et al., 2000; Marin et al., 2005). In the Bayesian PL mixture applications presented in Section 4 we exploited alternative relabeling algorithms, that perform an *ex-post* rearrangement of the raw MCMC drawings in order to obtain meaningful posterior estimates. These were implemented by means of the functions included in the recently released R package `label.switching` (Papastamoulis, 2016).

3.4. Determining the number of components

In the estimation procedures previously described the number G of groups is fixed *a priori*. Thus, after performing a separate inference on PL mixtures with a different number of components, a method for discriminating among the competing models is needed. In our applications we explored three types of alternative Bayesian criteria to address this issue: (i) *Deviance Information Criterion* (DIC) introduced by Spiegelhalter et al. (2002), (ii) *Bayesian*

Information Criterion-Monte Carlo (BICM) proposed by Raftery et al. (2007) and (iii) *Bayesian Predictive Information Criterion* (BPIC) described in Ando (2007). For an updated detailed review of Bayesian tools for model comparison see Gelman et al. (2014). We start from the general formula $\text{DIC} = \bar{D} + p_D$, where $\bar{D} = \mathbb{E}[D(\theta)|\underline{\pi}^{-1}]$ is the posterior expected deviance with $D(\theta) = -2\log L(\theta)$ and p_D represents the *effective number of parameters*. We consider two alternative DIC formulations corresponding to two alternative ways of conceiving p_D, i.e., $\text{DIC}_1 = \bar{D} + (\bar{D} - D(\hat{\theta}_{\text{MAP}}))$, based on the MAP estimate $\hat{\theta}_{\text{MAP}}$, and $\text{DIC}_2 = \bar{D} + \mathbb{VAR}[D(\theta)|\underline{\pi}^{-1}]/2$, suggested by Gelman et al. (2004). As shown in Raftery et al. (2007), DIC_2 coincides with AICM, that is, the Bayesian counterpart of AIC. We also use two versions of BICM, specifically: $\text{BICM}_1 = \bar{D} + \frac{\mathbb{VAR}[D(\theta)|\underline{\pi}^{-1}]}{2}(\log N - 1)$, which is based on the approximation of the MAP estimate from the MCMC sample (Raftery et al., 2007), and $\text{BICM}_2 = D(\hat{\theta}_{\text{MAP}}) + \frac{\mathbb{VAR}[D(\theta)|\underline{\pi}^{-1}]}{2}\log N$. Finally, since one aspect often debated on DIC is its tendency to overfit, due to the double usage of the observed data, we additionally employ two BPIC formulations obtained from DIC_1 and DIC_2 by doubling their penalty term p_D.

After fitting mixture models with alternative number of components, one can select a suitable number $\hat{G}$ of components by using a specific criterion and identifying the optimal mixture which minimizes that criterion. Since alternative criteria can lead to different choices, we will compare and discuss the possibly different selections of optimal models and provide some recommendation based on a simulation study.

3.5. Model assessment

Once the optimal PL mixture model has been selected, a comprehensive inferential analysis should also contemplate the adequacy of the estimated model in describing the observed data (Gelman et al., 1996). In this regard, we have focused on two important features of the ranking data $\underline{\pi}^{-1}$:

(i) the most-liked item frequency vector $\underline{r}(\underline{\pi}^{-1})$, whose generic entry $r_i(\underline{\pi}^{-1}) = \sum_{s=1}^{N} I_{[\pi_s^{-1}(1)=i]}$ counts how many times item i is ranked first;

(ii) the paired comparison frequency matrix $\tau(\underline{\pi}^{-1})$, whose generic entry

$$\tau_{ii'}(\underline{\pi}^{-1}) = \sum_{s=1}^{N}(1-(1-u_{si})(1-u_{si'}))I_{[\pi_s(i)<\pi_s(i')]} = \sum_{s=1}^{N}(u_{si}+u_{si'}-u_{si}u_{si'})I_{[\pi_s(i)<\pi_s(i')]}$$

counts the number of times that item i is preferred to item i'.

Within the Bayesian paradigm it is possible to generalize the classical goodness-of-fit statistic into a parameter-dependent quantity, referred to as *discrepancy variable* (Gelman et al., 1996), and perform a posterior predictive check of model goodness-of-fit. Let us denote with $\underline{\pi}_{\text{obs}}^{-1}$ the observed collection of partial orderings and with $\underline{\pi}_{\text{rep}}^{-1}$ a replicate random draw from the posterior predictive distribution under the specified model H. The *posterior predictive p-value* based on a generic discrepancy variable $X^2(\underline{\pi}^{-1};\theta)$ is defined as

$$p_B = \mathbf{P}\left(X^2(\underline{\pi}_{\text{rep}}^{-1};\theta) \geq X^2(\underline{\pi}_{\text{obs}}^{-1};\theta)\,\middle|\,\underline{\pi}_{\text{obs}}^{-1}, H\right). \tag{4}$$

Under correct model specification p_B is expected to be close to 0.5, whereas small values are deemed as an indication of model inadequacy. Here we considered 0.05 as critical threshold.

Indeed, as a first type of X^2 discrepancy measure we have considered

$$X^2_{(1)}(\underline{\pi}^{-1};\theta) = \sum_{i=1}^{K}\frac{\left(r_i(\underline{\pi}^{-1})-r_i^*(\theta)\right)^2}{r_i^*(\theta)},$$

where the symbol * indicates the theoretical frequency expected under PL mixture model with parameter $\theta = (\underline{p}, \underline{\omega})$. By following Yao and Böckenholt (1999), as a second discrepancy measure we have considered

$$X^2_{(2)}(\underline{\pi}^{-1};\theta) = \sum_{i<i'}\frac{(\tau_{ii'}(\underline{\pi}^{-1})-\tau_{ii'}^*(\theta))^2}{\tau_{ii'}^*(\theta)}.$$

Details on the computation of the posterior predictive p-values, denoted with $p_{B(d)}$ $d=1,2$ and corresponding to each discrepancy measure $X^2_{(d)}$, are reported in the Supplementary Material (SM).

Moreover, whenever partial ranking data show considerable proportions of strictly partial rankings with a different number of ranked items, one can further investigate model adequacy conditionally on the observed length of the partial orderings. This can help to verify possible violations of the underlying assumption that the subsets of rankers are identically distributed, such that their preference system is driven by the same mixture distribution on the support parameters. This check could reveal that one should better account for sample heterogeneity. To this aim, we have defined two other discrepancy measures, $\tilde{X}^2_{(1)}$ and $\tilde{X}^2_{(2)}$, which parallel the previous ones. The corresponding posterior predictive p-values, denoted with $\tilde{p}_{B(d)}$ $d = 1, 2$, allow to assess the homogeneity assumption of the strata of rankers characterized by different lengths of the expressed partial orderings. Details are reported in the SM.

4. Illustrative applications

We will apply our Bayesian model to simulated as well as two real data sets. We will verify its comparative performance with respect to some natural alternative methods, which can be expected to perform similarly, and highlight some possible advantages. We first provide some implementation details. Although the Bayesian approaches described in Section 3.2 and 3.3 permit to convey specific (subjective) prior knowledge on the parameters, in the following analyses we will rely upon weakly/noninformative prior densities with hyperparameters equal to $c_{gi} = 1$, $d_g = .001$ and $\alpha_g = 1$, in order to allow also for a direct comparison with the frequentist PL mixture developed by Gormley and Murphy (2006). For our parametric method, we first recorded the MAP estimate derived through the EM algorithm and subsequently employed it to initialize the GS. We run the MCMC algorithm for a total of 22000 iterations and discarded the first 2000 drawings as burn-in period. Moreover, the application of the alternative relabeling algorithms on the MCMC posterior samples revealed a good performance in removing the LS and returned very similar results in terms of adjusted estimates. Posterior means were used as final parameter estimates and those reported for the considered experiments were derived, specifically, with the application of the pivotal reordering algorithm (Marin et al., 2005).

In assessing the performance of our method we will focus also on the comparison with the BNPPLM, since it represents the most recent natural competitor to handle heterogeneity of partial ranking data and can be expected to perform similarly. In the BNPPLM analysis, we

Table 1: *Simulation study* - Percentages of top$-m$ partial orderings for the three censoring settings considered in the simulation study.

Censoring setting	m 1	2	3	4	5
A	0	2	4	10	84
B	5	15	15	20	45
C	5	20	20	25	30

employed the default setup for the hyperparameters described in Caron et al. (2014) and run their GS algorithm for 100000 iterations.

4.1. Simulation study

We considered a simulation plan with four different PL mixture population scenarios, where the true number G^* of components ranges from 1 to 4. Specifically, in Scenario $c \in \{1, \ldots, 4\}$ one has $G^* = c$. Under each scenario, we simulated 100 samples composed of $N = 1000$ complete orderings of $K = 6$ items. The values of the support parameters for each data set were randomly generated as $p_{gi} \overset{iid}{\sim} \text{Beta}(0.3, 0.3)$, where the U-shaped Beta density aims at guaranteeing a sufficient separation among the mixture components. Additionally, we assumed equal weights by setting $\omega_g = 1/G^*$ for all $g = 1, \ldots, G^*$ and $G^* = 1, \ldots, 4$. In order to perform the analysis on partial observations, a censoring was randomly induced on the complete orderings. In each scenario, we separately considered three censoring settings (A, B and C) for the random truncation of the complete data: the percentages of the number m of top ranked items are detailed in Table 1. In censoring setting A the percentages of partial orderings with the same number of ranked items were set equal to those observed in the CARCONF data considered in subsection 4.2. This yields approximately 16% of strictly partial orderings in each simulated sample. Censoring settings B and C are characterized by increasing proportions of truncation yielding, respectively, 55% and 70% of strictly partial observations. In this way we are able to thoroughly explore the effectiveness of our parametric framework and its sensitivity to differential presence of strictly partial rankings in the sample. Bayesian finite PL mixtures, with a number G of components ranging from 1 to 7, and the BNPPLM were fitted to all the artificial data sets for each population scenario and censoring setting. The comparison between the two models was based on the performance regarding the identification of the actual number G^* of groups in the four scenarios. In our Bayesian parametric PL mixture analysis the optimal number $\hat{G}$ of groups was identified by means of the alternative model selection criteria described in subsection 3.4. Tables ??, ?? and ?? show the distribution of $\hat{G}$ for the alternative criteria as well as that corresponding to the BNPPLM analysis obtained with the optimization method of Dahl (2006), as suggested by Caron et al. (2014). More briefly, Table 2 displays only the agreement rates (%). Regarding censoring setting A, in Scenario 1 BPIC_1, BPIC_2 and BICM_1 always recover the actual absence of heterogeneity (i.e. $G^* = 1$). On the other hand, this happens also for the frequentist approach employing BIC as well as for the BNPPLM. For the remaining population scenarios, BPIC_1 and DIC_1 perform from slightly to substantially better, especially in the case $G^* = 4$ where DIC_1 emerges with an agreement rate of 81%, followed by BPIC_1 with 77%. The gap with both the frequentist and nonparametric results is considerable. BIC exhibits an agreement rate equal to 66%, whereas for BNPPLM the rate is remarkably smaller. In fact, only for 50% of the data sets BNPPLM fitted $\hat{G} = G^*$ components. With the application of censoring setting B and C on the same synthetic data, we faced with the situation when most of the sequences to be analyzed are partial. With both censoring settings, for $G^* < 4$ the results associated to the Bayesian rules and BIC were found to be substantially robust with respect to censoring setting A and DIC_1 and BPIC_1 still differ in the best agreement rates. In the case $G^* = 4$, instead, the negative effect of the higher truncation percentage becomes more evident. In fact, we noted an overall worsening of the performance of all the selection methods, especially for the BNPPLM. Nonetheless, similarly to censoring setting A, DIC_1 exhibits the highest agreement rates (76% and 72%), confirming its sizable advantage over BIC (57% and 55%) and the BNPPLM (50% and 37%). Apparently, in almost all cases the BNPPLM approach yields the lowest agreement rates. Moreover, in the presence of heterogeneity ($G^* > 1$), BNPPLM is consistently associated with the greatest variability regarding the determination of the number $\hat{G}$ of groups, see the corresponding distributions in Table ??, ?? and ??. If on one hand the relatively worse performance of BNPPLM is partly due to the fact that data are simulated from a different generative model, on the other it highlights a possibly substantial difference between the two approaches. Overall, our

Table 2: *Simulation study* - Agreement rates (%) between true number G^* of PL components in the four population scenarios and the optimal number $\hat{G}$ of components identified, respectively, by the Bayesian PL mixture analysis via alternative model selection criteria and by the BNPPLM analysis via Dahl's procedure. Best agreement rates for each simulation scenario and censoring setting are highlighted in bold.

Censoring setting A								
G^*	DIC_1	DIC_2	$BPIC_1$	$BPIC_2$	$BICM_1$	$BICM_2$	BIC	BNPPLM
1	99	98	**100**	**100**	**100**	99	**100**	**100**
2	96	93	**98**	95	93	89	97	91
3	91	89	**94**	92	91	88	93	88
4	**81**	67	77	70	65	60	66	50
Censoring setting B								
G^*	DIC_1	DIC_2	$BPIC_1$	$BPIC_2$	$BICM_1$	$BICM_2$	BIC	BNPPLM
1	99	**100**	**100**	**100**	**100**	98	**100**	**100**
2	97	93	97	**98**	95	89	95	89
3	92	88	**94**	92	90	82	92	79
4	**76**	61	69	65	53	48	57	50
Censoring setting C								
G^*	DIC_1	DIC_2	$BPIC_1$	$BPIC_2$	$BICM_1$	$BICM_2$	BIC	BNPPLM
1	97	98	98	**100**	**100**	**100**	**100**	**100**
2	**97**	91	**97**	95	91	87	95	90
3	**95**	89	94	92	88	82	92	63
4	**72**	62	67	64	48	46	55	37

simulation study suggests to privilege the use of $BPIC_1$ and DIC_1 for the subsequent analysis, although with a higher occurrence of partial observations DIC_1 seems to be slightly preferable.

We conclude the simulation study by providing some evidence on the fitting measures presented in subsection 3.5. For succinctness, only results for the computation of $p_{B(d)}$ $(d = 1, 2)$ under the most critical population scenario with $G^* = 4$ components are shown. Boxplots of $p_{B(d)}$ values for all the parametric PL mixtures fitted to the 100 simulated data sets are reported in Figure **??** as a function of the number G of fitted components. They point out the effectiveness of the proposed diagnostic tools to highlight possible deficiencies of mis-specified PL mixture models. As expected, we observe an increasing trend of the posterior predictive p-values over $G < 4$, whereas for $G \geq 4$ they stabilize around the reference value 0.5.

4.2. The CARCONF data

Our second analysis concerns a marketing study described in Dabic and Hatzinger (2009), aimed at investigating customer preferences towards different car features. The car configurator

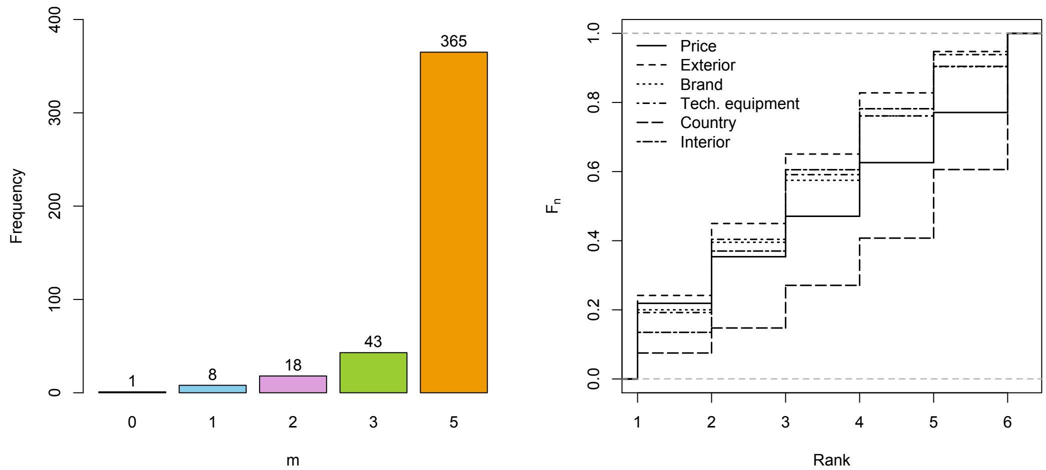

Figure 1: *CARCONF data* - Distribution of the number m of ranked items (left) and empirical c.d.f.'s of the marginal rank distributions of the car features (right).

(CARCONF) data set consists of $N = 435$ top orderings and is available in the R package `prefmod` (Hatzinger and Dittrich, 2012). Customers were asked to construct their car by using an online configurator system. The respondents were presented a set of $K = 6$ car modules to carry out their personal preferences, namely: 1=price, 2=exterior design, 3=brand, 4=technical equipment, 5=producing country and 6=interior design. The survey did not require a complete ranking elicitation, therefore the sample is composed of partial top orderings. The distribution of the varying number of ranked items is detailed in Figure 1 (left). Most of the customers (365 units, 84% of the sample) submitted a complete ordering of the car features. Most of the remaining customers submitted a strictly partial ordering by providing their top-4 favorite features. The vector (42,17,0,29,62,27) lists the number of missing responses for each item. Hence, all respondents assigned a rank to the brand, whereas the producing country is the one whose exact position is more frequently missing (62 occurrences corresponding to 89% of the total number of incomplete responses). The producing country is also associated with the lowest mean rank, as indicated by the fifth entry of the average rank vector $\bar{\pi} = (3.56, 2.88, 3.17, 3.11, 4.49, 3.20)$. The graphical inspection of the marginal rank distribution for each item, reported in the form of empirical c.d.f. in Figure 1 (right), provides additional details on the overall preferences. We note that the c.d.f. for the producing country is remarkably stochastically dominated by the other ones, matching the idea of a minor global interest in the car production country. Another important aspect to be highlighted is the presence of intersections among the c.d.f.'s, that can be interpreted as an empirical violation to the assumption of an underlying homogeneous PL, under which the rank distributions are instead expected to be *marginally stochastically ordered* (Marden, 1995). The observed behavior of the rank distributions could be explained with the existence of differential preference patterns in the sample. These patterns could be better captured by assuming an underlying group structure with an unknown number of groups, rather than with a basic homogeneous PL.

We estimated PL mixtures on the CARCONF data with a number of components varying from $G = 1$ to $G = 6$. Bayesian selection criteria and BIC are shown in Figure 2 (left). Numerical details are in Table **??**. BIC, as well as BICM_1 and BICM_2, does not recognize the existence of an underlying group structure despite the large sample size, whereas all versions of DIC and BPIC agree in selecting the 2-component PL mixture. The application of the BNPPLM to the CARCONF data set agrees with the MLE inference identifying a single PL component with vector of estimated support parameters equal to $\underline{\hat{p}} = (0.123, 0.231, 0.195, 0.193, 0.071, 0.187)$. In fact, the homogeneity conclusions of the *ad-hoc* criterion in Dahl (2006) matches with the degenerate distribution of the number of distinct support parameter vectors associated to each unit across the posterior simulations. This result conflicts somehow with our preliminary descriptive findings on the violation of the stochastic dominance among the marginal rank distributions.

For all the fitted Bayesian PL mixtures, posterior predictive p-values are reported in the last two columns of Table **??**. The posterior predictive p-value $p_{B(2)} = 0.505$ highlights a good fit in terms of the ability of the model to reproduce the bivariate features related to the pairwise comparisons, whereas $p_{B(1)} = 0.079$ reveals a possible deficiency of the model to recover the marginal probability of the most favorite item, although it is larger than the usual 0.05 critical threshold. On the other hand, we notice that $p_{B(1)} < 10^{-4}$ is well below for $G = 1$, supporting the need of a heterogeneous model.

Parameter estimates of the optimal 2-component PL mixture are displeyd in Figure 2 (center and right) and detailed in Table **??**. The selected mixture model suggests the presence of a major

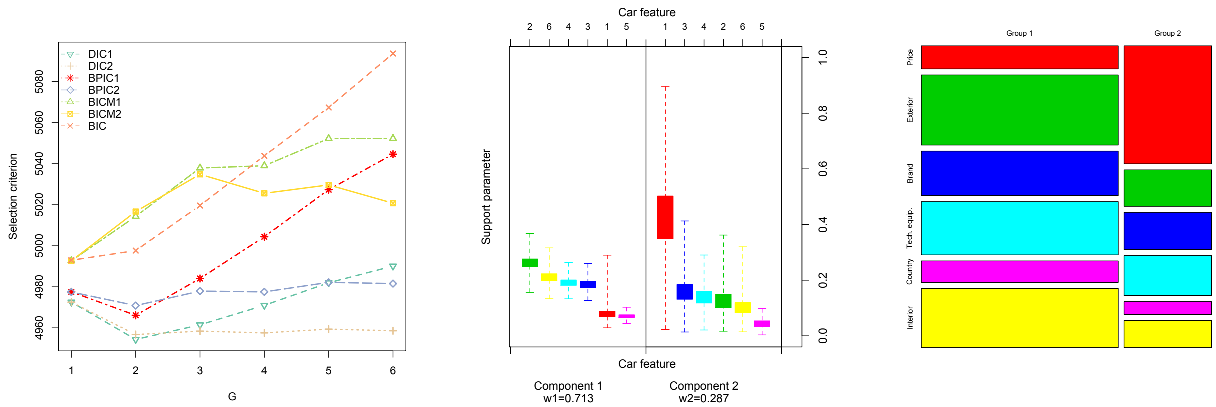

Figure 2: *CARCONF data* - Model selection criteria (left) for the Bayesian PL mixtures with a varying number G of components. For each selection criterion the optimal choice of the number of components corresponds to the minimum value. Boxplots (center) and mosaic plot (right) for the posterior samples of the support parameters of the optimal Bayesian 2-component PL mixture.

cluster ($\hat{\omega}_1 = 0.713$) comprised of customers mainly interested in aesthetic features, with greater support to the exterior ($\hat{p}_{12} = 0.263$) rather than interior design ($\hat{p}_{16} = 0.211$). The minor group ($\hat{\omega}_2 = 0.287$), instead, is characterized by a greater attention in the economic aspect represented by the price ($\hat{p}_{21} = 0.436$). Both groups share a minor interest in the production country ($\hat{p}_{15} = 0.071$ and $\hat{p}_{25} = 0.043$). These results seem to better accord with the typical preference patterns observed in the car market than the homogeneous scenario.

4.3. The APA data

Another interesting data set involving partial rankings is the popular 1980 American Psychological Association (APA) election data set. The entire APA data set with $N = 15449$ voters ranking a maximum of $K = 5$ candidates is available in the R package `ConsRank`. The majority of the ballots (9711, 63%) contain strictly partial orderings of the most favorite candidates and in most of them (5141, 33%) just a single favorite candidate is recorded. Some descriptive statistics are reported in Figure 3 and in the SM (Table **??** and **??**). A detailed explanation of the data collection and the corresponding voting system yielding the elected

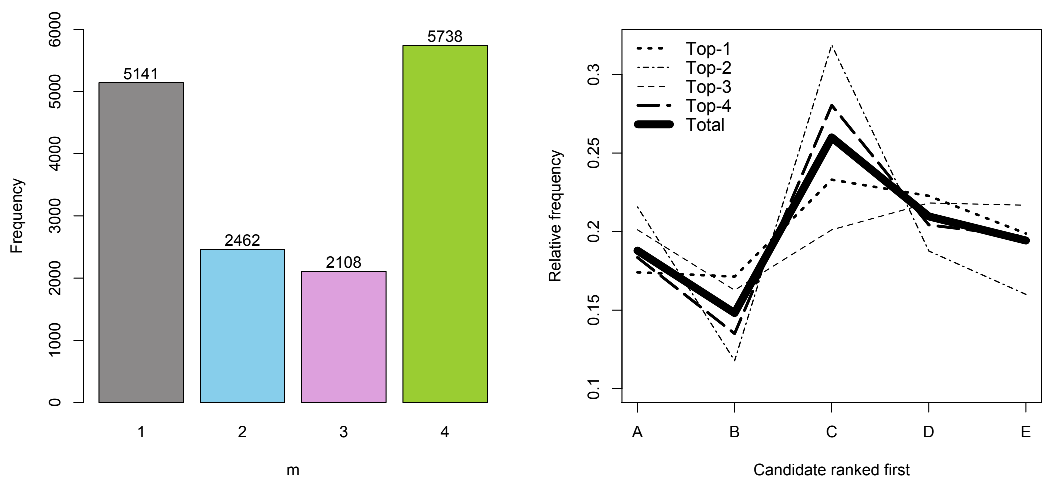

Figure 3: *APA election data* - Distribution of the number m of ranked items (left) and distribution of the candidate occupying the first position by number of ranked items (right).

candidate can be found in Diaconis (1987).

The popularity of these data is testified by the numerous attempts to provide an account of the complex heterogeneous structure of the ballots. From the pioneering descriptive analysis by Diaconis (1987), relying on the spectral group representation, to the most recent model-based approach by Jacques and Biernacki (2014), only few works proposed a probabilistic model for the whole set of 15449 partial orderings. Here we will show at what extent PL mixture models are able to provide an in-depth overall account of the underlying group structure.

We fitted Bayesian PL mixtures with $G = 1, \ldots, 12$ components to the whole APA data set. A relatively parsimonious PL mixture is selected by our parametric approach by using the Bayesian selection criteria displayed in Figure **??** (numerical values are reported in Table **??**). Indeed, BICM_1 and BICM_2 agree with BIC in selecting 5 components. However, as suggested by our simulation study, we privilege the use of BPIC_1 and DIC_1 which both agree in identifying $\hat{G} = 10$ groups. On the other hand, the alternative BNPPLM analysis adopting Dahl's procedure yields a partition of the electorate in 86 distinct clusters. Prior to illustrating the interpretation of the fitted components, we provide some new insights on model assessment. For the selected model, $p_{B(1)} = 0.471$ and $p_{B(2)} = 0.582$ do not highlight overall lack-of-fit, see Table **??**. However, the substantial presence of strictly partial orderings suggested a more specific check considering the conditional distributions of the same univariate and bivariate preference features within each subset $\underline{\pi}_m^{-1}$ of partial top orderings with the same length $m = 1, \ldots, 4$. It revealed that the best global model fitted to the whole set of ballots is unsuitable to describe the heterogeneity of these subsets. In fact, the corresponding $\tilde{p}_{B(1)} < 10^{-4}$ and $\tilde{p}_{B(2)} < 10^{-4}$ are well below the conventional 0.05 critical threshold and suggest to implement our PL mixture model separately on each subset, in order to provide a more appropriate account of the heterogeneity in the APA election data. We will then compare these results with those corresponding to our initial PL mixture analysis on the entire data set. Thus, we estimated the Bayesian PL mixture separately on top-1, top-2, top-3 and top-4 (full) orderings. Notice that on top-1 orderings only the PL mixture with $G = 1$ can be fully identified, since they correspond to ordinary multinomial data on K categories. Selection of the optimal number of components is displayed in Figure **??** (numerical values are reported in Table **??**). Indeed, if we analyze separately each subset and comment overall on all the resulting subgroups, we get a total of 1+2+ 3+7=13 clusters. We believe that these clusters provide a more appropriate representation of the heterogeneity in the APA election data. Unlike the overall model fitted to all the available ballots, for each model we get satisfactory fitting diagnostics (Table **??**). Now let us focus on the support parameter estimates of the different components fitted to each subset (Figure **??**). We notice that all the components exhibit distinctive modal orderings, apart from one component which shares the same modal pattern (C,A,B,E,D). Only three of them are recovered in the groups fitted to the whole data set (Figure **??**). Moreover, in none of the modal orderings of the components fitted with the separate analyses Candidate B is ranked first. Instead, in two out of ten components of the global model (corresponding to a total weight 0.16) Candidate B occupies the first position of the corresponding modal ordering with a relatively large estimated support parameter. This is, at a certain extent, surprising since Candidate B is that less frequently ranked first in all the subsets, see Figure 3 (right) and the corresponding Table **??**. Another interesting evidence from the separate analysis is that almost all components have Candidate C, D, or E in the first position of the modal orderings. The only exception, which provides maximum support to Candidate A, is found in a component fitted to the subset of full orderings. Such exceptional component, with estimated weight $\hat{\omega}_1 = 0.05$ in that subset, amounts to 1.9% of the entire data set. Additionally, if we aggregate the relative weights of all components resulting from the separate analysis which have Candidate C (the winner of the election) in the first position of the modal ordering, we get a total weight of 0.556. The analogue computation on the global mixture returns a total weight of 0.30. Notice, however, that both analyses provide a similar posterior mean vector of the support parameters, equal to $\underline{\hat{p}} = (0.189, 0.148, 0.259, 0.208, 0.196)$ in the global analysis and $\underline{\hat{p}} = (0.192, 0.148, 0.257, 0.206, 0.197)$ with the aggregation of the separate mixtures. Finally, by looking specifically at the results of the analysis on the 5738 top-4 (full) orderings, we can compare our findings with those of previous analyses. We found a larger number of components than Diaconis (1987) and Stern (1993), who both identified three clusters of voters, whereas Jacques and Biernacki (2014) reported the lowest BIC for ten components, although with a similar BIC corresponding to four components. Indeed, some vectors of support parameters characterizing our group structure well compare with the findings in Stern (1993), especially those for which Candidate C is in the first position of the modal ordering. Overall, our findings allow for a characterization of the groups in terms of a more marked preference for one or two candidates.

5. Concluding remarks and future work

We have investigated a Bayesian finite PL mixture for dealing with heterogeneous partially ranked data and described efficient algorithms to conduct posterior inference. Our proposal contemplates a data augmentation step with the latent group structure and allows for model-based classification of partial top orderings. It can be seen as a direct extension to the finite mixture context of the basic Bayesian PL introduced by Caron and Doucet (2012), aimed at identifying and characterizing possible groups of rankers with similar preferences/attitudes. On the other hand, it can be regarded as a Bayesian generalization of the PL mixture developed by Gormley and Murphy (2006), whose frequentist approach can be recovered as a by-product of the noninformative analysis. An important advantage over the MLE perspective lies in the possibility to straightforwardly address estimation uncertainty, without relying on large sample approximations and additional computational burden.

We have investigated the effectiveness of our estimation algorithms in a simulation study with multiple heterogeneity scenarios. In particular, we focused on the ability to recover the actual number of clusters of the generative mixture configuration. Our Bayesian parametric proposal provided a quite satisfactory performance, even when compared with the frequentist approach as well as with the Bayesian nonparametric alternative offered by the BNPPLM in Caron et al. (2014). Our simulations highlighted sometimes remarkable divergences in the final determination of the number of clusters, possibly due to the theoretically different notion of "group" behind the two Bayesian models. The analysis of two real experiments provided further evidence on the usefulness of our parametric model. For the CARCONF data the existence of a heterogeneous pattern of preferences emerged neither from the BNPPLM nor from the frequentist approach, whereas our proposal identified a 2-component PL mixture with two meaningful differential profiles. In general, estimating a smaller number of groups means that some preference patterns would not be recognized, leading to a less informative picture of the underlying preference system. On the other hand, the nonparametric method could prove itself more flexible to recover possible departures from the reference parametric ranking distribution by fitting a higher number of minor clusters to the sample. Summing up, both simulations and real data set analyses highlighted that our Bayesian finite PL mixture and the BNPPLM can lead to substantially different conclusions and, sometimes, our proposal could be preferred. This happens despite the fact that the nonparametric BNPPLM method could be regarded somehow as a generalization of the Bayesian finite PL mixture.

Additionally, this work provided some incremental findings on the performance of many alternative Bayesian selection criteria. Our investigation suggests, besides the most frequently adopted DIC_1, the use of BPIC_1. Also BIC performed well for smaller values of G^*. However, for larger value of G^* we confirm BIC's tendency to underestimate the true number of groups, as also pointed out in other mixture settings, see for example Celeux and Soromenho (1996); Lukočienė and Vermunt (2009) and Bulteel et al. (2013). In line with this evidence, under Scenario 4 no overestimation is present with BIC; on the other hand, BIC leads to underestimate the true number G^* of components for at least 30% of the simulated data sets in all the three censoring settings. Indeed, one could argue that, as a function of the sample size, the penalty term of BIC does not account for the varying rate of truncation, leading to a too severe penalization and, hence, to the selection of more parsimonious models. Conversely, with DIC_1 and BPIC_1 the effective number of parameters depends on the posterior deviance distribution, that inherently penalizes for the increasing parameterization and the higher censoring rate. For this reason, the two Bayesian criteria could be expected to return a more adaptive and suitable estimation of model complexity. Certainly, a more theoretical advancement is needed before a clear-cut conclusion on the most suitable criterion to adopt in the finite mixture framework, where regularity conditions facilitating the derivation of asymptotic results do not hold. Indeed, apart from few recent attempts (Miller and Harrison, 2013, 2014), in the nonparametric setting the asymptotic behavior is even less explored and understood.

We also made use of diagnostic devices to evaluate the fitting of our proposal via a posterior predictive check. Despite its practical relevance, the fitting performance is often neglected by practitioners, especially in the frequentist analysis of ranking data. Unlike previous applications in the partial ranking literature, we have also applied discrepancy measures accounting for the conditional distributions given the number of ranked items. These allow us to gain a more in-depth understanding of the adequacy of the PL parametric assumption in the whole data set.

A possible future development could be the Bayesian estimation of the mixture of Extended PL recently introduced by Mollica and Tardella (2014). One can extend model flexibility by exploiting the additional *reference order* parameter, representing the rank attribution order followed by the ranker to sequentially carry out his comparative judgment on the available items. Another interesting extension could be the introduction of extra information provided by individual and/or item-specific covariates. As revealed by previous applications (Gormley and Murphy, 2008, 2010), explanatory variables may fruitfully contribute to characterize choice patterns and support decisions for better capturing specific preference profiles or segments. Finally, the lack-of-fit due to the differential preference patterns underlying the different subsets of rankers who provide the same number of partially ranked items highlights the need of a more comprehensive model accounting for this type of observed heterogeneity.

References

Alvo, M. and Yu, P. L. (2014). *Statistical methods for ranking data*. Springer.

Ando, T. (2007). Bayesian predictive information criterion for the evaluation of hierarchical Bayesian and empirical Bayes models. *Biometrika*, 94(2):443–458.

Bulteel, K., Wilderjans, T. F., Tuerlinckx, F., and Ceulemans, E. (2013). CHull as an alternative to AIC and BIC in the context of mixtures of factor analyzers. *Behavior Research Methods*, 45(3):782–791.

Caron, F. and Doucet, A. (2012). Efficient Bayesian inference for Generalized Bradley-Terry models. *J. Comput. Graph. Statist.*, 21(1):174–196.

Caron, F., Teh, Y. W., and Murphy, T. B. (2012). Bayesian nonparametric Plackett-Luce models for the analysis of clustered ranked data. Technical Report 8143, Project-Team ALEA.

Caron, F., Teh, Y. W., and Murphy, T. B. (2014). Bayesian nonparametric Plackett-Luce models for the analysis of preferences for college degree programmes. *The Annals of Applied Statistics*, 8(2):1145–1181.

Celeux, G., Hurn, M., and Robert, C. P. (2000). Computational and inferential difficulties with mixture posterior distributions. *Journal of the American Statistical Association*, 95(451):957–970.

Celeux, G. and Soromenho, G. (1996). An entropy criterion for assessing the number of clusters in a mixture model. *Journal of Classification*, 13(2):195–212.

Dabic, M. and Hatzinger, R. (2009). Zielgruppenadaequate Ablaeufe in Konfigurationssystemen – Eine empirische Studie im Automobilmarkt – Partial Rankings. In Hatzinger, R., Dittrich, R., and Salzberger, T., editors, *Praeferenzanalyse mit R: Anwendungen aus Marketing, Behavioural Finance und Human Resource Management*. Facultas, Wien.

Dahl, D. B. (2006). Model-based clustering for expression data via a Dirichlet process mixture model. *Kim-Anh Do, Peter Müller & Marina Vannucci (Eds.)*, pages 201–218.

Diaconis, P. W. (1987). Spectral analysis for ranked data. Technical Report 282, Dept of Statistics, Stanford University.

Gelman, A., Carlin, J. B., Stern, H. S., and Rubin, D. B. (2004). *Bayesian data analysis*. Chapman & Hall/CRC, Boca Raton, FL, Second edition.

Gelman, A., Hwang, J., and Vehtari, A. (2014). Understanding predictive information criteria for Bayesian models. *Statistics and Computing*, 24(6):997–1016.

Gelman, A., Meng, X.-L., and Stern, H. (1996). Posterior predictive assessment of model fitness via realized discrepancies. *Statistica Sinica*, 6(4):733–760.

Gormley, I. C. and Murphy, T. B. (2006). Analysis of Irish third-level college applications data. *Journal of the Royal Statistical Society: Series A*, 169(2):361–379.

Gormley, I. C. and Murphy, T. B. (2008). A mixture of experts model for rank data with applications in election studies. *Ann. Appl. Stat.*, 2(4):1452–1477.

Gormley, I. C. and Murphy, T. B. (2009). A grade of membership model for rank data. *Bayesian Analysis*, 4(2):265–295.

Gormley, I. C. and Murphy, T. B. (2010). Clustering ranked preference data using sociodemographic covariates. In Hess, S. and Daly, A., editors, *Choice Modelling: The State-of-the-Art and the State-of-Practice: Proceedings from the Inaugural International Choice Modelling Conference*, pages 543–569. Emerald.

Guiver, J. and Snelson, E. (2009). Bayesian inference for Plackett-Luce ranking models. In Bottou, L. and Littman, M., editors, *Proceedings of the 26th International Conference on Machine Learning – ICML 2009*, pages 377–384. Omnipress.

Hatzinger, R. and Dittrich, R. (2012). prefmod: An R package for modeling preferences based on paired comparisons, rankings, or ratings. *Journal of Statistical Software*, 48(10):1–31.

Hunter, D. R. (2004). MM algorithms for Generalized Bradley-Terry models. *Ann. Statist.*, 32(1):384–406.

Jacques, J. and Biernacki, C. (2014). Model-based clustering for multivariate partial ranking data. *Journal of Statistical Planning and Inference*, 149:201–217.

Luce, R. D. (1959). *Individual choice behavior: A theoretical analysis*. John Wiley & Sons Inc.

Lukočienė, O. and Vermunt, J. K. (2009). Determining the number of components in mixture models for hierarchical data. In *Advances in data analysis, data handling and business intelligence*, pages 241–249. Springer.

Marden, J. I. (1995). *Analyzing and modeling rank data*, volume 64 of *Monographs on Statistics and Applied Probability*. Chapman & Hall.

Marin, J.-M., Mengersen, K., and Robert, C. P. (2005). Bayesian modelling and inference on mixtures of distributions. *Handbook of Statistics*, 25:459–507.

McCullagh, P., Yang, J., et al. (2008). How many clusters? *Bayesian Analysis*, 3(1):101–120.

Miller, J. W. and Harrison, M. T. (2013). A simple example of Dirichlet process mixture inconsistency for the number of components. In *Neural Information Processing Systems – NIPS 2013*, pages 199–206.

Miller, J. W. and Harrison, M. T. (2014). Inconsistency of Pitman-Yor process mixtures for the number of components. *The Journal of Machine Learning Research*, 15(1):3333–3370.

Mollica, C. and Tardella, L. (2014). Epitope profiling via mixture modeling of ranked data. *Statistics in Medicine*, 33(21):3738–3758.

Papastamoulis, P. (2016). label. switching: An R package for dealing with the Label Switching problem in MCMC outputs. *Journal of Statistical Software*, 69(1):1–24.

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

Raftery, A. E., Satagopan, Jaya, M., Newton, M. A., and Krivitsky, P. N. (2007). Bayesian statistics 8. In Bernardo, J., Bayarri, M., Berger, J., Dawid, A., Heckerman, D., Smith, A., and

West, M., editors, *Proceedings of the eighth Valencia International Meeting, June 2-6, 2006*, pages 371–416. Oxford University Press.

Spiegelhalter, D. J., Best, N. G., Carlin, B. P., and Van Der Linde, A. (2002). Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 64(4):583–639.

Stern, H. (1993). Probability models on rankings and the electoral process. In *Probability models and statistical analyses for ranking data*, volume 80 of *Lecture Notes in Statist.*, pages 173–195. Springer, New York.

Yao, G. and Böckenholt, U. (1999). Bayesian estimation of Thurstonian ranking models based on the Gibbs sampler. *British Journal of Mathematical and Statistical Psychology*, 52(1):79–92.

Yellott, John I., J. (1977). The relationship between Luce's choice axiom, Thurstone's theory of comparative judgment, and the double exponential distribution. *J. Mathematical Psychology*, 15(2):109–144.

June 1, 2016

BAYESIAN MIXTURE OF PLACKETT-LUCE MODELS FOR PARTIALLY RANKED DATA

Supplementary Material

Implementation details for model assessment: posterior predictive checks $p_{B(1)}$ and $p_{B(2)}$

We remind that the posterior predictive p-value represents the posterior probability that a parameter-dependent discrepancy measure $X^2_{(d)}(\underline{\pi}^{-1}_{\text{obs}};\theta)$, comparing actually observed frequencies and expected frequencies under the assumed model H, does not exceed the same discrepancy measure $X^2_{(d)}(\underline{\pi}^{-1}_{\text{rep}};\theta)$ evaluated on a replicated data set drawn from the same model, that is,

$$p_{B(d)} = \mathbf{P}\left(X^2_{(d)}(\underline{\pi}^{-1}_{\text{rep}};\theta) \geq X^2_{(d)}(\underline{\pi}^{-1}_{\text{obs}};\theta) \middle| \underline{\pi}^{-1}_{\text{obs}}, H \right).$$

The value $p_{B(d)}$ can be easily approximated by using the posterior simulations of the parameter vector θ and augmenting them with the corresponding draws of replicated data. For our first discrepancy measure $X^2_{(1)}(\underline{\pi}^{-1};\theta) = \sum_{i=1}^{K} \frac{\left(r_i(\underline{\pi}^{-1}) - r_i^*(\theta)\right)^2}{r_i^*(\theta)}$, the theoretical frequencies expected under PL mixture model with parameter $\theta = (\underline{p}, \underline{\omega})$ depend on the marginal overall support parameters $p_i = \sum_{g=1}^{G} \omega_g p_{gi}$ (for $i = 1, \ldots, K$) and are easily determined as follows

$$r_i^*(\theta) = N p_i.$$

For the discrepancy measure $X^2_{(2)}(\underline{\pi}^{-1};\theta) = \sum_{i<i'} \frac{(\tau_{ii'}(\underline{\pi}^{-1}) - \tau^*_{ii'}(\theta))^2}{\tau^*_{ii'}(\theta)}$, one can derive the expected paired comparison frequencies under PL mixture model as follows

$$\tau^*_{ii'}(\theta) = T_{ii'} \frac{p_i}{p_i + p_{i'}},$$

where $T_{ii'} = \tau_{ii'} + \tau_{i'i}$ indicates the total number of pairwise comparisons between item i and i'.

Implementation details for model assessment: posterior predictive checks $\tilde{p}_{B(1)}$ and $\tilde{p}_{B(2)}$

Let $m = 1, \ldots, K-1$ be the generic number of ranked items in a partial ordering of K items. We denote with $\underline{\pi}_m^{-1} = \{\pi_s^{-1} : n_s = m\}$ the subsample of N_m top-m orderings ($\sum_{m=1}^{K-1} N_m = N$). In order to assess the model adequacy about the homogeneity assumption on the conditional distributions given the same number m of ranked items, we define the discrepancy between each conditional distribution and the marginal distribution of the most-liked item by using the conditional frequencies $r_{i,m}$ as follows

$$\tilde{X}^2_{(1)}(\underline{\pi}^{-1};\theta) = \sum_{m=1}^{K-1}\sum_{i=1}^{K}\frac{(r_{i,m} - r^*_{i,m}(\theta))^2}{r^*_{i,m}(\theta)},$$

where $r_{i,m} = r_i(\underline{\pi}_m^{-1})$ and $r^*_{i,m}(\theta) = N_m p_i$. Similarly, when we aim at assessing homogeneity of the conditional pairwise comparison frequencies, we define

$$\tilde{X}^2_{(2)}(\underline{\pi}^{-1};\theta) = \sum_{m=1}^{K-1}\sum_{i<i'}\frac{(\tau_{ii',m} - \tau^*_{ii',m}(\theta))^2}{\tau^*_{ii',m}(\theta)},$$

where $\tau_{ii',m} = \tau_{ii'}(\underline{\pi}_m^{-1})$ and $\tau^*_{ii',m}(\theta) = T_{ii',m}\frac{p_i}{p_i + p_{i'}}$ with $T_{ii',m} = \tau_{ii',m} + \tau_{i'i,m}$. The computation of $\tilde{p}_{B(1)}$ and $\tilde{p}_{B(2)}$ follows the general formula (4) in the main paper by replacing the desired discrepancy measure.

Supplemental tables and figures

Table SM-1.

Simulation study (Censoring setting A) - Distribution (%) of the optimal number $\hat{G}$ of components identified, respectively, by the Bayesian PL mixture analysis via alternative model selection criteria and by the BNPPLM analysis via Dahl's procedure. In the simulation study 100 data sets with 1000 partial orderings of 6 items were generated from each PL mixture scenario with alternative true number G^* of components. Best agreement rates under each simulation scenario are highlighted in bold.

$G^* = 1$

$\hat{G}$	DIC_1	DIC_2	$BPIC_1$	$BPIC_2$	$BICM_1$	$BICM_2$	BIC	BNPPLM
1	99	98	**100**	**100**	**100**	99	**100**	**100**
2	1	2	0	0	0	1	0	0
3	0	0	0	0	0	0	0	0
4	0	0	0	0	0	0	0	0
5	0	0	0	0	0	0	0	0
6	0	0	0	0	0	0	0	0
7	0	0	0	0	0	0	0	0

$G^* = 2$

$\hat{G}$	DIC_1	DIC_2	$BPIC_1$	$BPIC_2$	$BICM_1$	$BICM_2$	BIC	BNPPLM
1	2	2	2	2	6	6	3	4
2	96	93	**98**	95	93	89	97	91
3	2	3	0	3	1	3	0	4
4	0	1	0	0	0	2	0	0
5	0	0	0	0	0	0	0	1
6	0	0	0	0	0	0	0	0
7	0	0	0	0	0	0	0	0

$G^* = 3$

$\hat{G}$	DIC_1	DIC_2	$BPIC_1$	$BPIC_2$	$BICM_1$	$BICM_2$	BIC	BNPPLM
1	0	0	0	0	0	0	0	0
2	4	4	5	6	8	8	7	6
3	91	89	**94**	92	91	88	93	88
4	5	5	1	2	1	3	0	5
5	0	1	0	0	0	1	0	1
6	0	0	0	0	0	0	0	0
7	0	1	0	0	0	0	0	0

$G^* = 4$

$\hat{G}$	DIC_1	DIC_2	$BPIC_1$	$BPIC_2$	$BICM_1$	$BICM_2$	BIC	BNPPLM
1	0	0	1	1	1	1	1	1
2	0	0	0	0	3	3	3	1
3	18	14	22	22	30	30	30	25
4	**81**	67	77	70	65	60	66	50
5	1	10	0	3	1	6	0	19
6	0	7	0	4	0	0	0	3
7	0	2	0	0	0	0	0	1

TABLE SM-2.

Simulation study (Censoring setting B) - Distribution (%) of the optimal number $\hat{G}$ of components identified, respectively, by the Bayesian PL mixture analysis via alternative model selection criteria and by the BNPPLM analysis via Dahl's procedure. In the simulation study 100 data sets with 1000 partial orderings of 6 items were generated from each PL mixture scenario with alternative true number G^* of components. Best agreement rates under each simulation scenario are highlighted in bold.

$G^* = 1$

$\hat{G}$	DIC_1	DIC_2	$BPIC_1$	$BPIC_2$	$BICM_1$	$BICM_2$	BIC	BNPPLM
1	99	**100**	**100**	**100**	**100**	98	**100**	**100**
2	1	0	0	0	0	2	0	0
3	0	0	0	0	0	0	0	0
4	0	0	0	0	0	0	0	0
5	0	0	0	0	0	0	0	0
6	0	0	0	0	0	0	0	0
7	0	0	0	0	0	0	0	0

$G^* = 2$

$\hat{G}$	DIC_1	DIC_2	$BPIC_1$	$BPIC_2$	$BICM_1$	$BICM_2$	BIC	BNPPLM
1	2	2	3	2	5	5	5	4
2	97	93	97	**98**	95	89	95	89
3	1	3	0	0	0	5	0	6
4	0	2	0	0	0	1	0	1
5	0	0	0	0	0	0	0	0
6	0	0	0	0	0	0	0	0
7	0	0	0	0	0	0	0	0

$G^* = 3$

$\hat{G}$	DIC_1	DIC_2	$BPIC_1$	$BPIC_2$	$BICM_1$	$BICM_2$	BIC	BNPPLM
1	0	0	0	0	0	0	0	0
2	5	5	5	6	10	11	8	7
3	92	88	**94**	92	90	82	92	79
4	3	5	1	2	0	7	0	11
5	0	2	0	0	0	0	0	2
6	0	0	0	0	0	0	0	0
7	0	0	0	0	0	0	0	1

$G^* = 4$

$\hat{G}$	DIC_1	DIC_2	$BPIC_1$	$BPIC_2$	$BICM_1$	$BICM_2$	BIC	BNPPLM
1	1	1	1	1	1	1	1	1
2	0	0	0	1	6	6	5	2
3	21	21	30	30	39	39	37	19
4	**76**	61	69	65	53	48	57	50
5	2	9	0	3	1	6	0	20
6	0	7	0	0	0	0	0	5
7	0	1	0	0	0	0	0	2
8	0	0	0	0	0	0	0	1

Table SM-3.

Simulation study (Censoring setting C) - Distribution (%) of the optimal number $\hat{G}$ of components identified, respectively, by the Bayesian PL mixture analysis via alternative model selection criteria and by the BNPPLM analysis via Dahl's procedure. In the simulation study 100 data sets with 1000 partial orderings of 6 items were generated from each PL mixture scenario with alternative true number G^* of components. Best agreement rates under each simulation scenario are highlighted in bold.

$G^* = 1$

$\hat{G}$	DIC_1	DIC_2	$BPIC_1$	$BPIC_2$	$BICM_1$	$BICM_2$	BIC	BNPPLM
1	97	98	98	**100**	**100**	**100**	**100**	**100**
2	3	1	2	0	0	0	0	0
3	0	0	0	0	0	0	0	0
4	0	0	0	0	0	0	0	0
5	0	0	0	0	0	0	0	0
6	0	1	0	0	0	0	0	0
7	0	0	0	0	0	0	0	0

$G^* = 2$

$\hat{G}$	DIC_1	DIC_2	$BPIC_1$	$BPIC_2$	$BICM_1$	$BICM_2$	BIC	BNPPLM
1	2	2	3	3	8	8	5	5
2	**97**	91	**97**	95	91	87	95	90
3	1	6	0	2	1	5	0	5
4	0	1	0	0	0	0	0	0
5	0	0	0	0	0	0	0	0
6	0	0	0	0	0	0	0	0
7	0	0	0	0	0	0	0	0

$G^* = 3$

$\hat{G}$	DIC_1	DIC_2	$BPIC_1$	$BPIC_2$	$BICM_1$	$BICM_2$	BIC	BNPPLM
1	0	0	0	0	0	0	0	0
2	4	5	6	6	12	11	8	8
3	**95**	89	94	92	88	82	92	63
4	1	5	0	2	0	5	0	22
5	0	1	0	0	0	1	0	6
6	0	0	0	0	0	0	0	1
7	0	0	0	0	0	1	0	0

$G^* = 4$

$\hat{G}$	DIC_1	DIC_2	$BPIC_1$	$BPIC_2$	$BICM_1$	$BICM_2$	BIC	BNPPLM
1	0	0	1	1	2	2	1	1
2	1	1	0	1	9	9	6	2
3	25	23	32	30	41	39	38	22
4	**72**	62	67	64	48	46	55	37
5	2	7	0	3	0	1	0	16
6	0	6	0	1	0	0	0	13
7	0	1	0	0	0	3	0	6
8	0	0	0	0	0	0	0	3

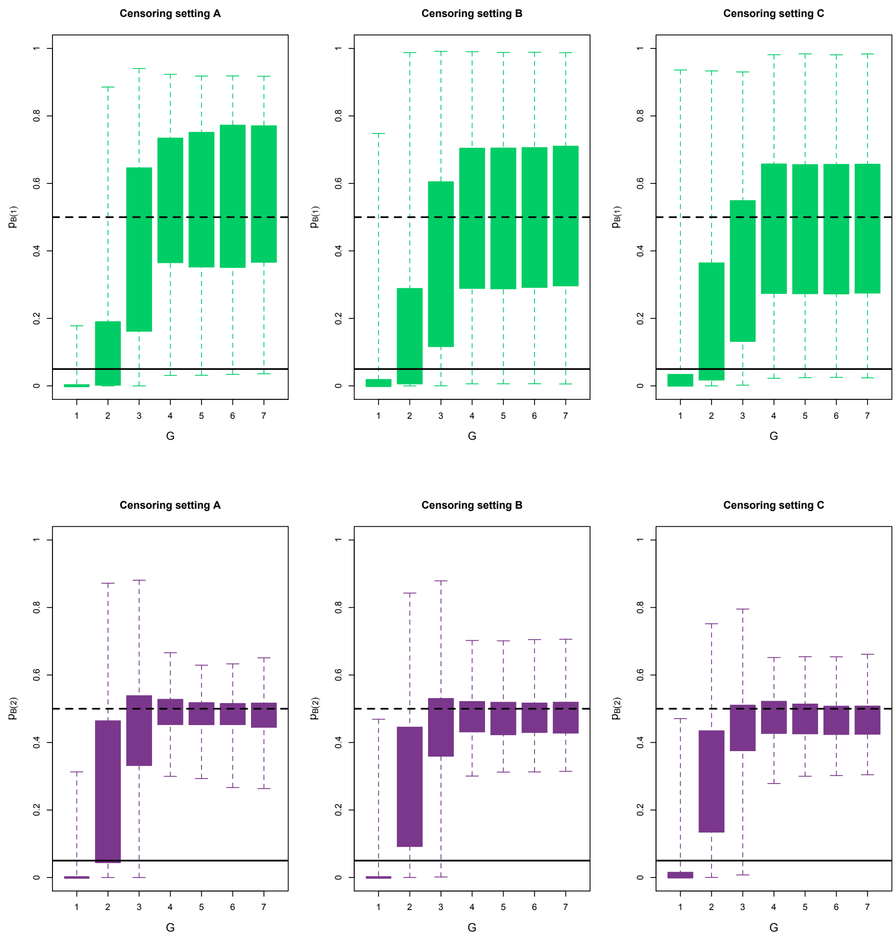

Figure SM-1.

Simulation study (Scenario 4) - Model assessment criteria for the Bayesian PL mixtures fitted to the 100 data sets simulated from the population scenario with $G^* = 4$ groups as a function of the number G of fitted components. The solid and dashed line represent, respectively, the critical threshold 0.05 and the reference value 0.5 expected under correct model specification.

TABLE SM-4.

CARCONF data (435 full and partial orderings) - Model selection criteria and posterior predictive p-values for the Bayesian PL mixtures with a varying number G of components. For each selection criterion the optimal choice of the number of components corresponds to the minimum value (in bold).

G	DIC_1	DIC_2	BPIC_1	BPIC_2	BICM_1	BICM_2	BIC	$p_{B(1)}$	$p_{B(2)}$
1	5288.34	5288.29	5293.32	5293.24	**5308.44**	**5308.39**	**5308.74**	0.000	0.247
2	**5268.73**	**5268.90**	**5280.15**	**5280.48**	5316.09	5316.25	5312.73	0.079	0.505
3	5278.45	5273.38	5301.99	5291.84	5348.62	5343.55	5334.66	0.092	0.515
4	5289.34	5272.67	5324.82	5291.47	5349.29	5332.61	5358.12	0.103	0.508
5	5295.06	5273.46	5336.93	5293.75	5356.14	5334.55	5387.49	0.107	0.516
6	5305.01	5274.43	5357.28	5296.12	5362.83	5332.25	5413.11	0.122	0.518

TABLE SM-5.

CARCONF data (435 full and partial orderings) - Posterior means of the parameters and component-specific modal orderings of the optimal Bayesian 2-component PL mixture. Posterior standard deviations are shown in parentheses.

g	$\hat{\omega}_g$		$\hat{\sigma}_g^{-1}$	$\hat{p}_{g1}$		$\hat{p}_{g2}$		$\hat{p}_{g3}$		$\hat{p}_{g4}$		$\hat{p}_{g5}$		$\hat{p}_{g6}$	
1	.713	(.10)	(2,6,4,3,1,5)	.079	(.02)	.263	(.02)	.185	(.02)	.191	(.01)	.071	(.01)	.211	(.02)
2	.287	(.10)	(1,3,4,2,6,5)	.436	(.13)	.124	(.04)	.157	(.05)	.138	(.03)	.043	(.02)	.101	(.03)

TABLE SM-6.

APA election data (15449 full and partial orderings) - Percentage of voters who assign position t to Candidate i (upper panel) and average rank vector (lower panel).

	Candidate				
Rank	A	B	C	D	E
1	18.8	14.8	26.0	21.0	19.4
2	27.7	17.7	16.9	16.9	20.7
3	23.6	24.1	14.0	18.6	19.7
4	17.5	24.7	18.3	20.3	19.3
5	14.8	18.4	23.1	23.4	20.3
$\bar{\pi}$	2.37	2.66	2.34	2.51	2.47

Table SM-7.

APA election data (15449 full and partial orderings) - Percentage of voters who rank Candidate i in the first position conditionally on the number m of ranked candidates.

	Candidate				
m	A	B	C	D	E
1	17.4	17.1	23.3	22.3	19.9
2	21.6	11.8	31.9	18.8	16.0
3	20.1	16.3	20.1	21.8	21.7
4	18.4	13.5	28.0	20.4	19.7

Table SM-8.

APA election data (15449 full and partial orderings) - Model selection criteria and posterior predictive p-values for the Bayesian PL mixtures with a varying number G of components. For each selection criterion the optimal choice of the number of components corresponds to the minimum value (in bold).

Full + Partial orderings									
G	DIC_1	DIC_2	BPIC_1	BPIC_2	BICM_1	BICM_2	BIC	$p_{B(1)}$	$p_{B(2)}$
1	103204.59	103204.65	103208.57	103208.71	103235.66	103235.73	103235.19	0.000	0.000
2	100772.97	100772.87	100781.63	100781.44	100838.40	100838.30	100842.44	0.000	0.610
3	100591.84	100591.89	100603.00	100603.10	100677.58	100677.62	100704.56	0.004	0.493
4	100436.20	100445.60	100443.54	100462.34	100573.58	100582.98	100604.78	0.298	0.411
5	100396.98	100401.44	100413.46	100422.39	**100561.56**	**100566.03**	**100595.51**	0.343	0.523
6	100360.59	100369.50	100377.16	100394.99	100564.33	100573.24	100607.17	0.375	0.503
7	100336.08	100344.30	100361.35	100377.81	100600.44	100608.66	100613.47	0.233	0.536
8	100341.12	100347.07	100381.82	100393.71	100703.71	100709.65	100635.88	0.390	0.492
9	100341.55	100350.84	100390.59	100409.18	100796.84	100806.13	100667.85	0.426	0.531
10	**100317.68**	100346.31	**100358.35**	100415.63	100876.22	100904.86	100708.95	0.471	0.528
11	100321.54	**100314.49**	100376.02	**100361.92**	100677.10	100670.05	100733.43	0.382	0.531
12	100340.38	100349.80	100408.73	100427.57	100944.37	100953.79	100772.76	0.440	0.521

Table SM-9.

APA election data (separate analysis for subsets of ballots with the same number of ranked candidates) - Model selection criteria and posterior predictive p-values for the Bayesian PL mixtures with a varying number G of components. For each selection criterion the optimal choice of the number of components corresponds to the minimum value (in bold).

Top-2 orderings

G	DIC_1	DIC_2	$BPIC_1$	$BPIC_2$	$BICM_1$	$BICM_2$	BIC	$p_{B(1)}$	$p_{B(2)}$
1	14255.36	14255.24	14259.33	14259.09	14277.59	14277.47	14278.66	0.000	0.049
2	13427.79	13427.99	**13436.89**	**13437.29**	**13481.98**	**13482.18**	**13479.88**	0.295	0.502
3	13427.70	13426.86	13443.00	13441.33	13510.92	13510.08	13506.41	0.379	0.530
4	13434.10	13427.37	13458.77	13445.32	13531.62	13524.90	13533.12	0.410	0.530
5	13433.24	13425.80	13458.63	13443.74	13530.04	13522.60	13569.87	0.443	0.531
6	13431.30	13426.03	13455.70	13445.16	13537.16	13531.89	13608.96	0.455	0.527
7	13429.54	13425.14	13453.09	13444.29	13536.39	13531.99	13647.93	0.463	0.522
8	13429.62	13425.23	13453.22	13444.45	13536.84	13532.45	13686.96	0.472	0.526
9	13428.02	**13423.42**	13450.81	13441.60	13529.06	13524.45	13726.03	0.478	0.525
10	13427.65	13424.41	13450.28	13443.79	13537.01	13533.76	13765.02	0.479	0.522
11	13427.58	13423.66	13450.16	13442.32	13532.07	13528.15	13804.08	0.476	0.525
12	**13426.88**	13423.49	13449.11	13442.32	13532.91	13529.51	13843.13	0.478	0.513

Top-3 orderings

G	DIC_1	DIC_2	$BPIC_1$	$BPIC_2$	$BICM_1$	$BICM_2$	BIC	$p_{B(1)}$	$p_{B(2)}$
1	17147.88	17147.84	17151.92	17151.83	17170.41	17170.37	17170.43	0.000	0.180
2	16732.01	16733.01	16741.62	16743.63	**16793.05**	**16794.05**	**16781.65**	0.262	0.468
3	**16717.06**	16717.76	**16731.79**	**16733.19**	16805.00	16805.69	16794.74	0.278	0.529
4	16717.27	16719.43	16742.81	16747.13	16876.03	16878.19	16811.62	0.440	0.530
5	16718.74	16720.82	16752.55	16756.70	16923.69	16925.77	16834.81	0.479	0.523
6	16723.28	16713.24	16762.76	16742.70	16879.76	16869.73	16866.26	0.483	0.511
7	16725.94	16716.31	16769.12	16749.85	16905.96	16896.33	16899.80	0.477	0.516
8	16726.43	16715.82	16771.67	16750.46	16911.66	16901.06	16934.43	0.468	0.509
9	16729.16	16715.99	16776.54	16750.20	16909.40	16896.23	16971.16	0.497	0.509
10	16735.15	16716.82	16788.59	16751.92	16915.26	16896.92	17003.31	0.491	0.511
11	16736.19	16718.88	16790.71	16756.09	16929.25	16911.94	17040.45	0.478	0.507
12	16735.28	**16710.82**	16790.19	16741.29	16883.04	16858.58	17077.01	0.497	0.505

Top-4 (full) orderings

G	DIC_1	DIC_2	$BPIC_1$	$BPIC_2$	$BICM_1$	$BICM_2$	BIC	$p_{B(1)}$	$p_{B(2)}$
1	54812.60	54812.60	54816.60	54816.61	54839.29	54839.30	54839.21	0.000	0.000
2	53696.16	53695.65	53705.49	53704.48	53754.42	53753.91	53755.38	0.000	0.639
3	53576.87	53574.99	53591.48	53587.73	53659.78	53657.91	53668.81	0.043	0.655
4	53477.70	53477.33	53494.38	53493.63	53585.83	53585.46	**53608.79**	0.179	0.478
5	53458.70	53454.30	53479.35	53470.54	**53562.39**	**53557.98**	53625.13	0.183	0.470
6	53433.51	53439.25	53460.29	53471.77	53655.67	53661.42	53630.95	0.462	0.479
7	**53412.08**	53437.63	**53445.60**	53496.70	53830.73	53856.28	53639.31	0.503	0.512
8	53420.25	**53416.43**	53464.62	**53456.97**	53686.23	53682.40	53669.06	0.440	0.436
9	53449.15	53499.96	53512.83	53614.45	54261.90	54312.71	53702.59	0.413	0.489
10	53415.29	53443.76	53466.78	53523.72	53975.90	54004.37	53736.39	0.481	0.553
11	53437.74	53446.27	53503.70	53520.77	53942.07	53950.61	53773.17	0.403	0.462
12	53424.85	53438.45	53488.02	53515.22	53949.36	53962.97	53809.15	0.455	0.519

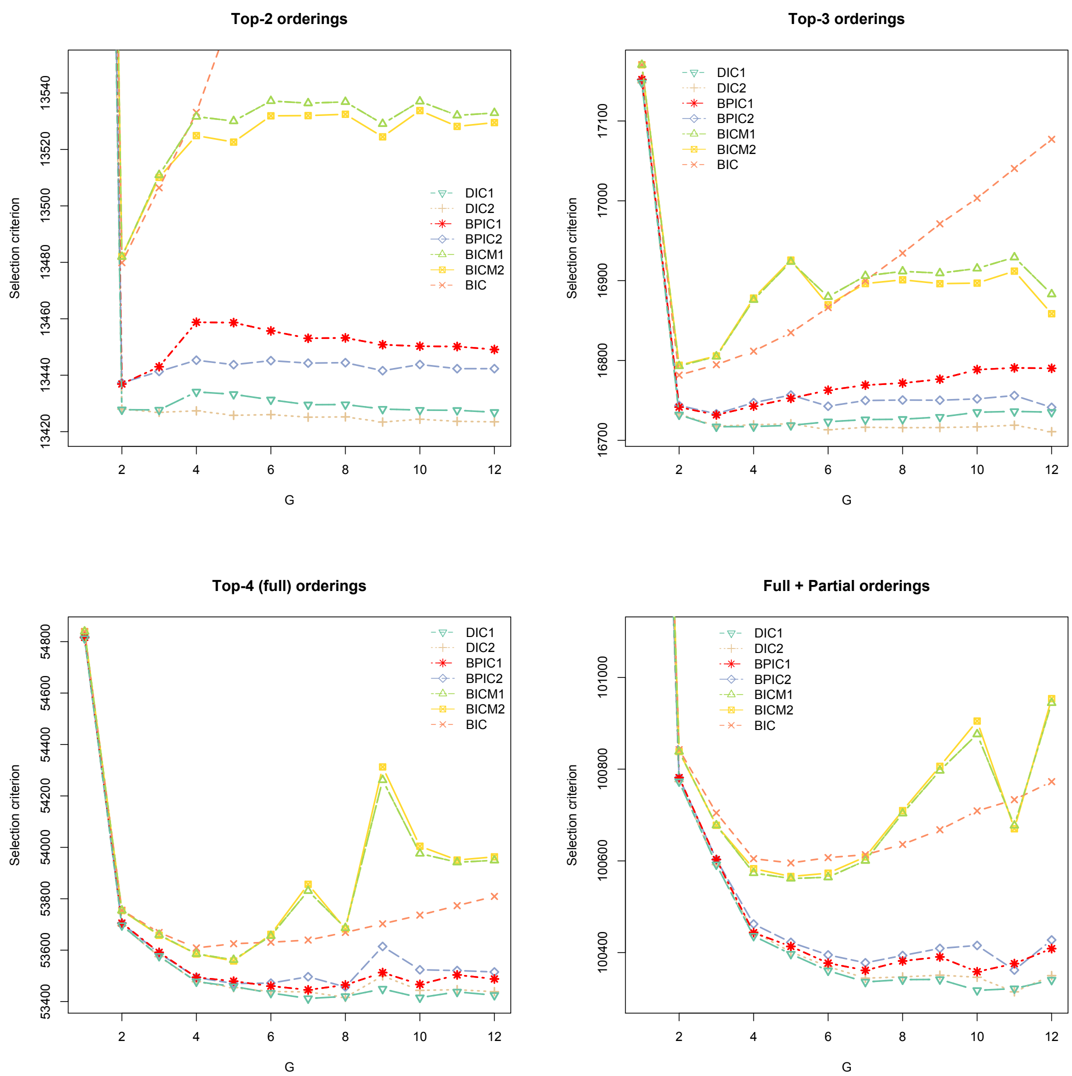

Figure SM-2.

APA election data (global and separate analysis for subsets of ballots with the same number of ranked candidates) - Model selection criteria for the Bayesian PL mixtures with a varying number G of components. For each selection criterion the optimal choice of the number of components corresponds to the minimum value.

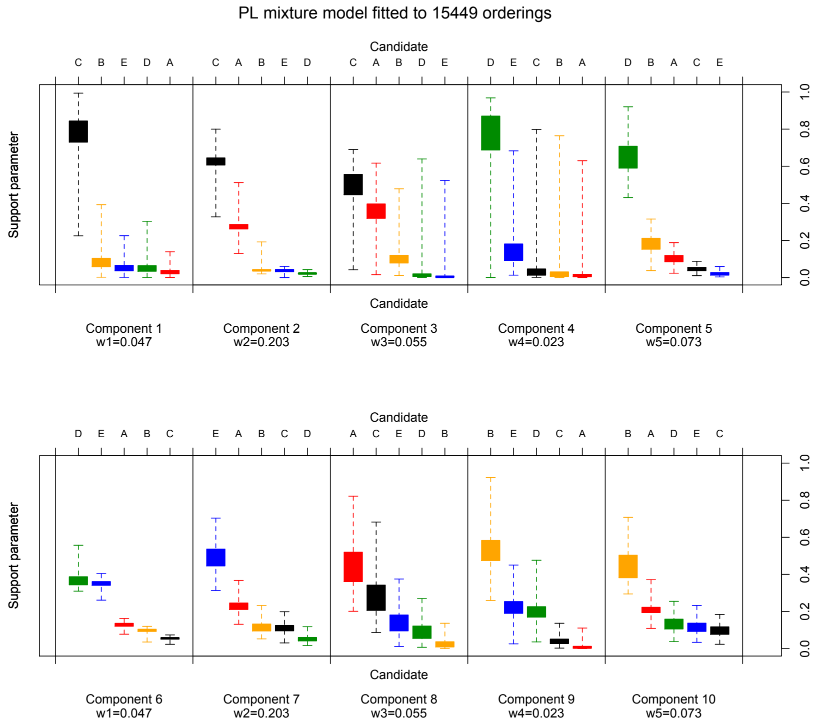

FIGURE SM-3.

APA election data (global analysis) - Optimal Bayesian 10-component PL mixture model fitted to the entire data set.

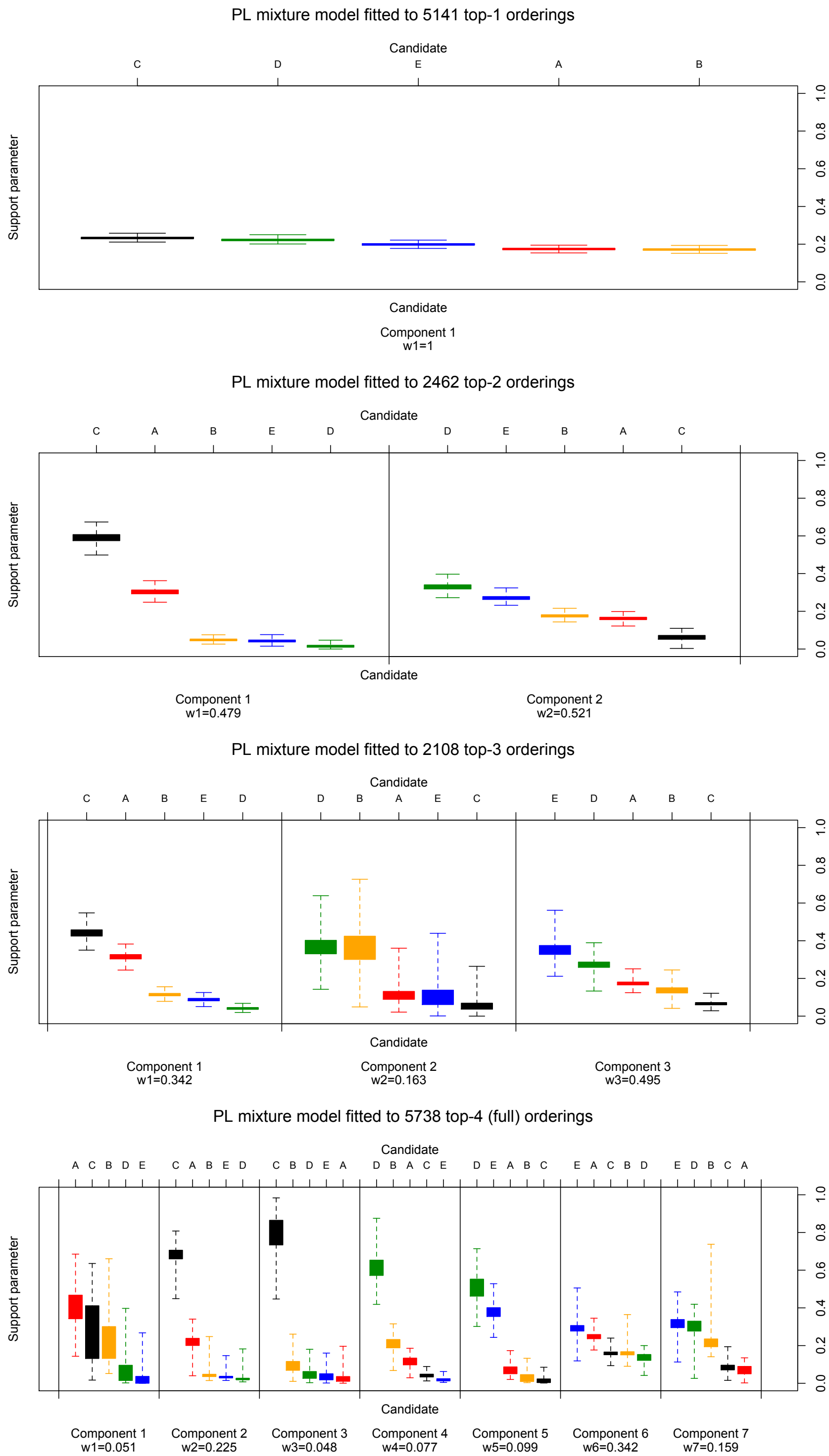

Figure SM-4.

APA election data (separate analysis) - Optimal Bayesian PL mixtures fitted to subsets of partial orderings with the same number m of ranked items.