



An Imprecise Label Ranking Method for Heterogeneous Data

Tathagata Basu, Sébastien Destercke, Benjamin Quost

► To cite this version:

Tathagata Basu, Sébastien Destercke, Benjamin Quost. An Imprecise Label Ranking Method for Heterogeneous Data. 10th International Conference on Soft Methods in Probability and Statistics (SMPS 2022), Sep 2022, Valladolid, Spain. pp.32-39, 10.1007/978-3-031-15509-3_5 . hal-03946248

HAL Id: hal-03946248

<https://hal.science/hal-03946248>

Submitted on 19 Jan 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



An Imprecise Label Ranking Method for Heterogeneous Data

Tathagata Basu, Sébastien Destercke, Benjamin Quost

► To cite this version:

Tathagata Basu, Sébastien Destercke, Benjamin Quost. An Imprecise Label Ranking Method for Heterogeneous Data. International Conference on Soft Methods in Probability and Statistics, Sep 2022, Valladolid (Spain), Spain. pp.32-39, 10.1007/978-3-031-15509-3_5 . hal-03943571

HAL Id: hal-03943571

<https://hal.science/hal-03943571>

Submitted on 17 Jan 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

An Imprecise Label Ranking Method for Heterogeneous Data

Tathagata Basu, Sébastien Destercke and Benjamin Quost

Abstract Learning to rank is an important problem in many sectors ranging from social sciences to artificial intelligence. However, it remains a rather difficult task to perform. Therefore, in some cases, it is preferable to perform cautious inference. For this purpose, we look into the possibility of an imprecise probabilistic approach for the Plackett-Luce model, a popular probabilistic model for label ranking. We aim at extending current Bayesian inference techniques for the Plackett-Luce model to an imprecise probabilistic setting so that we can deal with heterogeneous data by means of cautious mixture modelling. To achieve this, we perform a robust Bayesian analysis over a set of imprecise Dirichlet priors, which allows us to perform cautious label ranking. Finally, we use a synthetic dataset to illustrate our imprecise estimation method.

1 Introduction

Ranking objects is an important problem in many areas, such as social sciences, stock markets, e-commerce, etc. Sometimes, such rankings proceed from pairwise comparisons between the objects: one such treatment can be found in the model defined by Bradley and Terry (1952). This model naturally extends to the Plackett-Luce model in the case of multiple comparisons, as suggested by Plackett (1975) and Luce (1959). Several frequentist and Bayesian estimation methods have been developed based on these models.

Once the ranking model has been constructed, it can be used to estimate an optimal ranking between the objects. However, estimation of this ranking requires the data to be homogeneous, i.e. the objects are compared by a sub-

Tathagata Basu (✉) · Sébastien Destercke · Benjamin Quost
UMR CNRS 7253 Heudiasyc, Université de Technologie de Compiègne
Compiègne, France. e-mail: {name.surname}@hds.utc.fr

population of rankers which are assumed to be consistent with each other. In reality this might not be the case as we may gather this ranking data from different sources making the data heterogeneous. In such cases, mixtures of ranking models allow us to capture the sample information efficiently and also opens up the possibility of predicting a ranking after observing the ranker.

One of the first works on heterogeneous data was done by Gormley and Murphy (2006) where they suggested a mixture model using the Plackett-Luce model. Later, Caron et al. (2014) suggested a Bayesian alternative using a Dirichlet process model, where infinitely many models are assumed to be present in the mixture. A similar idea involving a finite mixture was proposed by Mollica and Tardella (2016) for partially ranked data. Recently, Adam et al. (2020) proposed an imprecise probabilistic approach for the Plackett-Luce model where imprecise estimation was carried out using likelihood cuts.

In this paper, we discuss the notion of robust Bayesian analysis for the Plackett-Luce model. Section 2 presents our approach, very similar to that of Mollica and Tardella (2016). In our approach we overcome the difficulty of obtaining closed forms for the imprecise estimates by using a non-linear optimiser for certain parameter estimates. Section 3 illustrates our approach on a synthetic dataset. Section 4 concludes the paper along with a discussion on future works.

2 The Plackett-Luce Model

The Plackett-Luce model (Plackett (1975)) is a simple and intuitive probabilistic model which gives us a probability for any observed ranking of p objects. Each object is associated with a strength parameter λ , which determines its probability of being preferred over others when drawing a sequence of objects. This model gives the probability of n independent rankings as

$$P(X \mid \lambda) = \prod_{i=1}^n \prod_{j=1}^{p_i-1} \frac{\lambda_{x_{ij}}}{\sum_{m=j}^{p_i} \lambda_{x_{im}}}. \quad (1)$$

where $p_i \leq p$ is the number of objects in the i -th ranking, $\lambda := (\lambda_1, \dots, \lambda_p)$ is the vector of strength parameters, and $X := [x_{ij}]$ is the $n \times p$ matrix containing the rankings (i.e., x_{ij} is the rank of the j th object or participant in the i th observed ranking, with $i = 1, \dots, n$ and $j = 1, \dots, p_i$). Note that Eq. (1) may be called the Plackett-Luce distribution because of its probabilistic formulation.

Example 1 Table 1 displays $n = 2$ rankings observed over $p = 4$ different objects ‘A’, ‘B’, ‘C’ and ‘D’; where A and B have rank 4 and 1 in the first observed ranking ($x_{12} = 1$ and $x_{11} = 4$). The probability of these data is given by Eq. (2):

	Obj. 1	Obj. 2	Obj. 3	Obj. 4
Ranking 1	B	D	C	A
Ranking 2	B	A	C	–

Table 1 Toy example with $n = 2$ rankings of $p = 4$ objects

$$P(X | \lambda) = \left[\frac{\lambda_B}{\lambda_B + \lambda_D + \lambda_C + \lambda_A} \cdot \frac{\lambda_D}{\lambda_D + \lambda_C + \lambda_A} \cdot \frac{\lambda_C}{\lambda_C + \lambda_A} \right] \cdot \left[\frac{\lambda_B}{\lambda_B + \lambda_A + \lambda_C} \cdot \frac{\lambda_A}{\lambda_A + \lambda_C} \right]. \quad (2)$$

We aim to estimate the strength parameters which maximise this probability.

2.1 Hierarchical model

We follow Mollica and Tardella (2016) to construct our hierarchical mixture model. Though, we interpret a partial ordering involving $K \leq p$ objects as a top- K ordering, assuming the remaining objects to be absent from the model. This also simplifies the expression of the PL distribution, which is beneficial for faster computation.

Assuming a total of G components in the mixture, the model can be written as

$$X_i | \lambda, \omega \sim \sum_{g=1}^G \omega_g \text{PL}(X_i | \lambda_g), \quad \text{PL}(X_i | \lambda_g) = \prod_{j=1}^{p_i-1} \frac{\lambda_{g,x_{ij}}}{\sum_{m=j}^{p_i} \lambda_{g,x_{im}}}.$$

We associate each observation with a unique latent membership indicator z_i , which follows a categorical distribution

$$z_i | \omega \sim \text{Cat}(\omega_1, \dots, \omega_G),$$

with ω_g being the weight of the g -th mixture component. For a full Bayesian treatment, we assign a set of imprecise Dirichlet priors on these weights:

$$\omega | s, \alpha \sim \text{Dir}(s; \alpha_1, \dots, \alpha_G),$$

where $\underline{\alpha}_g \leq \alpha_g \leq \bar{\alpha}_g$ for $g = 1, \dots, G$, with in addition $\sum_g \alpha_g = 1$ and $s > 0$.

For the data augmentation process, we follow Mollica and Tardella (2016) and use exponentially distributed variables y_{ij} so that

$$y_{ij} \stackrel{\text{ind}}{\sim} \text{Exp} \left(\prod_{g=1}^G \left(\sum_{m=j}^{p_i} \lambda_{g,x_{im}} \right)^{z_{ig}} \right).$$

Finally, we specify the strength parameters λ using gamma priors:

$$\lambda_{gk} \sim \text{Gamma}(a_{gk}, b_k),$$

with $\underline{a}_{gk} \leq a_{gk} \leq \bar{a}_{gk}$, for $1 \leq g \leq G$ and $1 \leq k \leq p$.

2.2 Parameter Estimation

To discuss parameter estimation, we first look at the following iterative steps. These formulas can be obtained by adding log-prior components to the complete data log-likelihood given by Gormley and Murphy (2006). Since details are provided in this latter reference, we omit them here.

$$\begin{aligned} \hat{z}_{ig}^{(t+1)} &= \frac{\hat{\omega}_g^{(t)} \text{PL}(X_i | \hat{\lambda}_g^{(t)})}{\sum_g \hat{\omega}_g^{(t)} \text{PL}(X_i | \hat{\lambda}_g^{(t)})}, \quad \hat{\omega}_g^{(t+1)} = \frac{s \alpha_g - 1 + \sum_i \hat{z}_{ig}^{(t+1)}}{s - G + N}, \\ \hat{\lambda}_{gk}^{(t+1)} &= \frac{a_{gk} - 1 + \sum_i \hat{z}_{ig}^{(t+1)} u_{ik}}{b_g + \sum_i \hat{z}_{ig}^{(t+1)} \sum_j \frac{\delta_{ijk}}{\sum_{m=j}^{p_i} \hat{\lambda}_{g, x_{im}}^{(t)}}}, \quad (3) \end{aligned}$$

where $u_{ik} = \mathbb{I}_{k \in \{x_{i1}, \dots, x_{i(p_i-1)}\}}$ and $\delta_{ijk} = \mathbb{I}_{k \in \{x_{ij}, \dots, x_{ip_i}\}}$. Note that, to start this iterative process we need suitable initial guesses for λ_g and ω_g .

2.2.1 Imprecise Estimates

In order to compute imprecise estimates, we need to calculate the bounds of the parameters in Eq. (3) over the sets of all possible values of $\alpha := (\alpha_1, \dots, \alpha_G)$. From the iterative formula of the posterior membership probability (\hat{z}_{ig}), we have

$$\begin{aligned} \hat{z}_{ig}^{(t+1)} &= \frac{\hat{\omega}_g^{(t)} \text{PL}(X_i | \hat{\lambda}_g^{(t)})}{\sum_g \hat{\omega}_g^{(t)} \text{PL}(X_i | \hat{\lambda}_g^{(t)})} = \frac{1}{1 + \frac{\sum_{g' \neq g} \hat{\omega}_{g'}^{(t)} \text{PL}(X_i | \hat{\lambda}_{g'}^{(t)})}{\hat{\omega}_g^{(t)} \text{PL}(X_i | \hat{\lambda}_g^{(t)})}} \\ &\geq \frac{1}{1 + \frac{\sum_{g' \neq g} \max_{\hat{\omega}_{g'}, \hat{\lambda}_{g'}^{(t)}} \left\{ \hat{\omega}_{g'}^{(t)} \text{PL}(X_i | \hat{\lambda}_{g'}^{(t)}) \right\}}{\min_{\hat{\omega}_g^{(t)}, \hat{\lambda}_g^{(t)}} \left\{ \hat{\omega}_g^{(t)} \text{PL}(X_i | \hat{\lambda}_g^{(t)}) \right\}}}. \end{aligned}$$

Now, let

$$\underline{\text{PL}}(X_i | \hat{\lambda}_g^{(t)}) := \min_{\hat{\lambda}_g^{(t)}} \left\{ \text{PL}(X_i | \hat{\lambda}_g^{(t)}) \right\}, \quad \overline{\text{PL}}(X_i | \hat{\lambda}_g^{(t)}) := \max_{\hat{\lambda}_g^{(t)}} \left\{ \text{PL}(X_i | \hat{\lambda}_g^{(t)}) \right\}$$

such that $\hat{\lambda}_{gk}^{(t)} \leq \hat{\lambda}_{gk}^{(t)} \leq \bar{\lambda}_{gk}^{(t)}$ for $1 \leq k \leq p$. Then, the lower bound of the posterior membership probability is given by:

$$\hat{z}_{ig} = \frac{\hat{\omega}_g^{(t)} \underline{\text{PL}}(X_i | \hat{\lambda}_g^{(t)})}{\hat{\omega}_g^{(t)} \underline{\text{PL}}(X_i | \hat{\lambda}_g^{(t)}) + \sum_{g' \neq g} \bar{\omega}_{g'}^{(t)} \overline{\text{PL}}(X_i | \hat{\lambda}_{g'}^{(t)})}, \quad (4)$$

and similarly the upper bound is given by:

$$\bar{z}_{ig} = \frac{\bar{\omega}_g^{(t)} \overline{\text{PL}}(X_i | \hat{\lambda}_g^{(t)})}{\bar{\omega}_g^{(t)} \overline{\text{PL}}(X_i | \hat{\lambda}_g^{(t)}) + \sum_{g' \neq g} \hat{\omega}_{g'}^{(t)} \underline{\text{PL}}(X_i | \hat{\lambda}_{g'}^{(t)})}, \quad (5)$$

where the lower and upper bounds $\hat{\omega}_g^{(t)}$ and $\bar{\omega}_g^{(t)}$ on the mixture weights $\hat{\omega}_g^{(t)}$ is obtained from Eq. (3), so that

$$\hat{\omega}_g^{(t)} = \frac{s\alpha_g - 1 + \sum_i \hat{z}_{ig}^{(t)}}{s - G + N} \quad \text{and} \quad \bar{\omega}_g^{(t)} = \frac{s\bar{\alpha}_g - 1 + \sum_i \bar{z}_{ig}^{(t)}}{s - G + N}. \quad (6)$$

Even though the update equations for these parameter bounds can be derived easily, computing these bounds is difficult and we need to employ an optimiser to compute $\underline{\text{PL}}(X_i | \hat{\lambda}_g^{(t)})$ and $\overline{\text{PL}}(X_i | \hat{\lambda}_g^{(t)})$

We face similar issues for the strength parameters as well, as $\hat{\lambda}_{gk}^{(t+1)}$ are not monotone with respect to $\hat{z}_{ig}^{(t+1)}$'s. As a result, in order to compute $\hat{\lambda}_{gk}^{(t+1)}$ and $\bar{\lambda}_{gk}^{(t+1)}$; we need to solve the following multiobjective optimisation problems:

$$\min \left(\hat{\lambda}_{1k}^{(t+1)}, \hat{\lambda}_{2k}^{(t+1)}, \dots, \hat{\lambda}_{Gk}^{(t+1)} \right) \quad \text{and} \quad \max \left(\hat{\lambda}_{1k}^{(t+1)}, \hat{\lambda}_{2k}^{(t+1)}, \dots, \hat{\lambda}_{Gk}^{(t+1)} \right), \quad (7)$$

such that $\hat{z}_{ig} \leq \hat{z}_{ig} \leq \bar{z}_{ig}$, for $1 \leq g \leq G$ and $1 \leq i \leq N$.

3 Illustration

We illustrate our approach using a synthetic dataset, which allows us to assess the performance of our method in estimating the strength parameters. We consider a set of $p = 8$ objects, together with two different sets of strengths λ : in one case, $(\lambda_{1,1}, \lambda_{1,2}, \dots, \lambda_{1,8}) = (8, 7, \dots, 1)$, and in the other case $(\lambda_{2,1}, \lambda_{2,2}, \dots, \lambda_{2,8}) = (1, 2, \dots, 8)$. These strengths correspond to two ranking processes which order the objects in an exactly opposite way.

We use these parameters to randomly generate our ranking dataset using the generative model suggested by Caron and Doucet (2012). Then, we pick 60% samples from the first set, and the remaining 40% from the second one. Note that, this is a randomised generative process and in the estimation the orderings of the components may not be estimated based on this generation.

Results

To show our results, we consider two different settings. In the first case, we consider the weights of the Dirichlet model to be precise and equal. In the second case, we define imprecise initial weights, such that $0.5 \leq \alpha_1 \leq 0.7$ and $0.3 \leq \alpha_2 \leq 0.5$. In both cases, we set our initial estimates for a_{gk} , so that they satisfy the orderings of λ_{gk} estimated from a precise estimation process. This initialisation step is crucial to avoid $\hat{\lambda}_g^{(t)}$ taking extreme values, which would be problematic for the convergence of the iterative algorithm. Moreover, in each iteration, we need to enforce orderings of these $\hat{\lambda}_{gk}^{(t)}$ estimates whilst computing $\underline{\text{PL}}(X_i | \hat{\lambda}_g^{(t)})$ and $\overline{\text{PL}}(X_i | \hat{\lambda}_g^{(t)})$ using the optimiser. Nevertheless, to ensure robustness, we take a large interval for each a_{gk} for capturing the imprecision in the data.

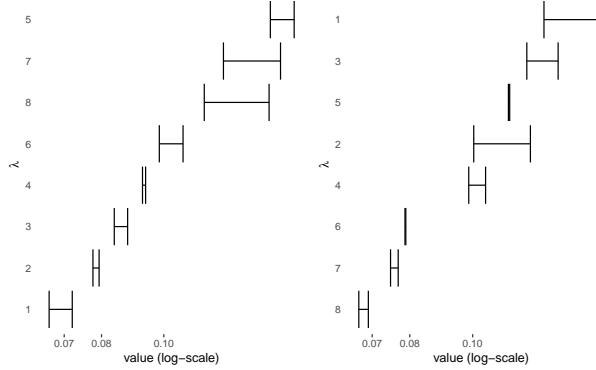


Fig. 1 Strength parameter estimates for $\alpha_1 = \alpha_2 = 0.5$ and $s = 5$.

For precise and equal initial weights, we get estimated mixture weights $\omega_1 \in [0.36; 0.63]$ and $\omega_2 \in [0.37; 0.64]$. The corresponding strengths are displayed in Fig. 1 (first component on left and the second one on right). We notice that the strength parameters follow a strict ordering for the first component, which is not the case for the second one¹: in this latter case, we notice that ‘2’ and ‘5’ are not comparable, which leads to a partial ordering.

¹ We consider, object i is preferred to object j if $\underline{\lambda}_i > \underline{\lambda}_j$ and $\overline{\lambda}_i > \overline{\lambda}_j$.

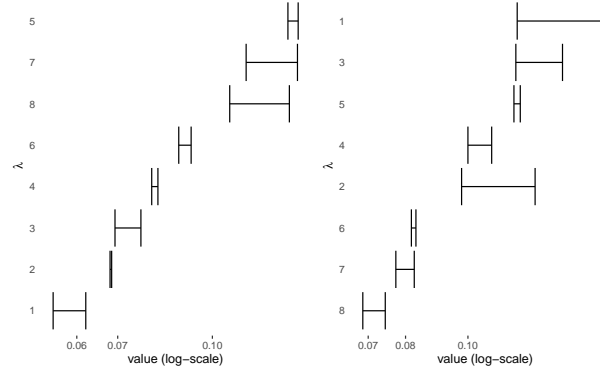


Fig. 2 Strength parameter estimates for $0.3 \leq \alpha_1 \leq 0.5$; $0.5 \leq \alpha_2 \leq 0.7$ and $s = 5$.

We notice a similar outcome for imprecise initial weights (Fig. 2). In this case, we obtain mixture weight estimates $\omega_1 \in [0.21; 0.62]$ and $\omega_2 \in [0.38; 0.78]$. We can observe from the figure that the intervals tend to be larger than the intervals in the first case, confirming the initial imprecise prior. Furthermore, we see that the estimated ranking from the first component is the same as in the precise initialisation case (see Fig. 1). However, the estimates are slightly different for the second component: then, ‘2’ is not comparable to both ‘4’ and ‘5’, though ‘5’ is strictly preferred to ‘4’.

We also notice that our choice of ordered imprecise weights gives us better estimates for the mixture weights, which is not the case for starting with equal weights. This happens as for the equal weights, the imprecision in the data is only captured through our ignorance about the size of the mixture components whereas for the ordered weights, our model tends to assign membership probabilities more efficiently and can also capture the imprecision in rankings.

4 Conclusion

In this article, we investigate inferring a probabilistic ranking model from heterogeneous data: we assume the data to come from several sub-populations, each of which can be associated with a Plackett-Luce ranking model. We propose a robust Bayesian approach to estimate the strength parameters of the resulting mixture of PL models. We notice that estimating the model is computationally expensive, since it requires to repeatedly calculate the bounds of each PL distribution — which involves solving nonlinear optimisation problems.

We illustrate our method using a synthetic dataset, in order to study its efficiency in inferring a mixture of rankings. The experiments show that im-

precise initial mixture weights tend to produce wider intervals for the strength estimates, which reflects our methods ability to showcase the inherent imprecision in scarce data. This is extremely important for cautious ranking as we may want to abstain from ordering in certain scenarios.

A major issue remaining to be addressed is the high computation time of the estimation procedure. This motivates us to find tight approximate bounds to ensure monotonicity. Hopefully, this will help us to reduce the computation cost significantly, while allowing us to investigate problems with a high number of mixture components.

Acknowledgements This research was funded by the project PreServe (ANR Grant ANR-18-CE23-0008).

References

- Adam L, Van Camp A, Destercke S, Quost B (2020) Inferring from an Imprecise Plackett–Luce Model: Application to Label Ranking. In: Springer (ed) 14th International Conference on Scalable Uncertainty Management (SUM 2020), Bolzano, Italy, Lecture Notes in Computer science, vol 12322, pp 98–112
- Bradley RA, Terry ME (1952) Rank Analysis of Incomplete Block Designs: I. The Method of Paired Comparisons. *Biometrika* 39(3/4):324–345
- Caron F, Doucet A (2012) Efficient Bayesian Inference for Generalized Bradley–Terry Models. *Journal of Computational and Graphical Statistics* 21(1):174–196
- Caron F, Teh YW, Murphy TB (2014) Bayesian nonparametric Plackett–Luce models for the analysis of preferences for college degree programmes. *The Annals of Applied Statistics* 8(2):1145 – 1181, DOI 10.1214/14-AOAS717
- Gormley IC, Murphy TB (2006) Analysis of Irish third-level college applications data. *Journal of the Royal Statistical Society: Series A (Statistics in Society)* 169(2):361–379, DOI <https://doi.org/10.1111/j.1467-985X.2006.00412.x>
- Luce RD (1959) *Individual Choice Behavior: A Theoretical analysis*. Wiley, New York, NY, USA
- Mollica C, Tardella L (2016) Bayesian Plackett–Luce mixture models for partially ranked data. *Psychometrika* 82(2):442–458, DOI 10.1007/s11336-016-9530-0
- Plackett RL (1975) The Analysis of Permutations. *Journal of the Royal Statistical Society: Series C (Applied Statistics)* 24(2):193–202, DOI <https://doi.org/10.2307/2346567>