



HAL
open science

Conjugate prior in the Mallows model with Spearman distance

Marta Crispino, Isadora Antoniano -Villalobos

► **To cite this version:**

Marta Crispino, Isadora Antoniano -Villalobos. Conjugate prior in the Mallows model with Spearman distance. 2018 ISBA World Meeting, Jun 2018, Edinburgh, United Kingdom. hal-01973165

HAL Id: hal-01973165

<https://hal.science/hal-01973165>

Submitted on 8 Jan 2019

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.

Introduction

► **Ranking and comparing items:** crucial for collecting information about preferences in many areas: e.g. marketing, business, politics, genetics.

► **Mallows model:** powerful model for rankings

► **Question:** How to include **expert opinions** into the analysis? How to elicit a meaningful **prior on the consensus** ranking of a population?

1. The Mallows model with Spearman distance

The **Mallows model** (Mallows 1957) is a class of non-uniform distributions for $\mathbf{R} \in \mathcal{P}_n$, the set of n -dimensional permutations, of the form

$$P(\mathbf{R}|\theta, \rho) = \frac{e^{-\theta d(\mathbf{R}, \rho)}}{Z(\theta)}$$

► $\rho \in \mathcal{P}_n$: consensus ranking, location parameter

► $\theta \geq 0$: precision parameter

► $d(\cdot, \cdot)$: right-invariant (Diaconis 1988) distance

► $Z(\theta) = \sum_{\mathbf{r} \in \mathcal{P}_n} e^{-\theta d(\mathbf{r}, \mathbf{1}_n)}$: partition function (independent of ρ because of right-invariance of $d(\cdot, \cdot)$)

Spearman distance: the squared l_2 norm on \mathcal{P}_n

$$d_S(\rho, \sigma) = \sum_{i=1}^n (\rho_i - \sigma_i)^2, \quad \rho, \sigma \in \mathcal{P}_n$$

Consider n items, ranked by N assessors.

Denote by $\mathbf{R}_j = (R_{1j}, R_{2j}, \dots, R_{nj}) \in \mathcal{P}_n$, the ranking of user j , $j = 1, \dots, N$.

Under the Mallows model with Spearman distance (MMS), given θ , the likelihood can be written as

$$P(\mathbf{R}_1, \dots, \mathbf{R}_N | \rho) \propto \exp\left(2\theta N \sum_{i=1}^n \rho_i \bar{R}_i\right)$$

where $\bar{R}_i = \frac{1}{N} \sum_{j=1}^N R_{ij}$, $i = 1, \dots, n$, is the sample average of the i -th rank.

2. Sufficient statistics and mle

The sufficient statistic for $\rho = (\rho_1, \dots, \rho_n)$, when θ is known, is $\bar{\mathbf{R}} = (\bar{R}_1, \dots, \bar{R}_n)$.

Proposition 1. Let $\mathbf{R}_1, \dots, \mathbf{R}_N | \rho, \theta \sim \text{Mall}(\theta, \rho)$, and define the vector of sample ranks $\bar{\mathbf{R}}$ as above. Assume $\bar{R}_i \neq \bar{R}_j$, for each $i \neq j$, and denote by $\mathbf{Y}(\bar{\mathbf{R}}) = [Y_1(\bar{\mathbf{R}}), \dots, Y_n(\bar{\mathbf{R}})] \in \mathcal{P}_n$ the rank vector of $\bar{\mathbf{R}}$, i.e. $Y_i(\bar{\mathbf{R}}) = Y_i = \sum_{h=1}^n \mathbb{1}(\bar{R}_h \leq \bar{R}_i)$, $i = 1, \dots, n$. Then the unique mle of ρ is

$$\rho_{mle} = \operatorname{argmax}_{\rho \in \mathcal{P}_n} \sum_{i=1}^n \rho_i \bar{R}_i = \mathbf{Y}(\bar{\mathbf{R}}).$$

Notice that, in general, $\bar{\mathbf{R}} \notin \mathcal{P}_n$. However $\bar{\mathbf{R}}$ lives in the permutohedron of order n .

Definition 1. The **permutohedron** of order n , $\mathbb{P}\mathcal{P}_n$, is an $(n-1)$ -dimensional polytope embedded in an n -dimensional space, the vertices of which are formed by permuting the coordinates of the vector $(1, 2, 3, \dots, n)$. Equivalently, it is the convex hull of the points $\rho \in \mathcal{P}_n$, the set of n -dim permutations.

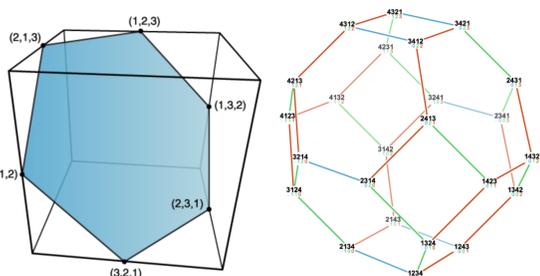


Figure 1: Left: The permutohedron of order 3 is a regular hexagon, filling a cross-section of a $2 \times 2 \times 2$ cube; Right: The permutohedron of order 4 is a truncated octahedron.

3. The Bayesian Mallows

Vitelli et al. (2018) proposed a framework for performing **Bayesian inference on the Mallows** model.

They assume that ρ is a priori uniformly distributed, $\rho \sim \text{Unif}(\mathcal{P}_n)$.

Contribution: the objective prior in the sense of Villa & Walker (2015) is the uniform prior density on the space of permutations $\rho \sim \text{Unif}(\mathcal{P}_n)$.

Can we go further?

How to put an informative prior on ρ ?

A: θ given: conjugate prior for ρ

B: θ not given: conjugate conditioned on θ + clever prior on $\theta \rightarrow$ MH sampling scheme to approximate the posterior

3.A Conjugate prior for ρ (given θ)

Proposition 2. Keeping θ fixed, the conjugate prior for $\rho \in \mathcal{P}_n$ is

$$\pi(\rho | \rho_0, \theta N_0) = \frac{\exp\left[-\theta N_0 \sum_{i=1}^n (\rho_i - \rho_{0i})^2\right]}{Z^*(\theta N_0, \rho_0)},$$

where $\rho_0 \in \mathbb{P}\mathcal{P}_n$, and $N_0 \in \mathbb{N}$. We call $\pi(\cdot | \rho_0, \theta N_0)$ the extended Mallows density: it is a Mallows model where the consensus ρ_0 belongs to $\mathbb{P}\mathcal{P}_n$.

Posterior consensus: weighted average of prior parameter and observed mean (recall Diaconis et al. 1979):

$$\rho_N = \frac{N}{N_0 + N} \bar{\mathbf{R}} + \frac{N_0}{N_0 + N} \rho_0 \in \mathbb{P}\mathcal{P}_n$$

$$\theta_N = \theta(N_0 + N) \in [0, \infty)$$

Then N_0 can be interpreted as an equivalent sample size (recall the Gaussian).

Remark 1. When $N_0 = 0$, the conjugate prior reduces to the uniform, for all ρ_0 . When $\rho_0 = (\frac{n+1}{2}, \dots, \frac{n+1}{2})$ the conjugate prior reduces to the uniform, for all θ_0 .

3.A.bis Toy example 1

Sample $N = 30$ rankings from the MMS with $\rho = (3, 1, 2)$ and $\theta = 0.18$. Conjugate prior with $\rho_0 = (1, 2, 3)$, and varying $\theta_0 = \theta N_0$.

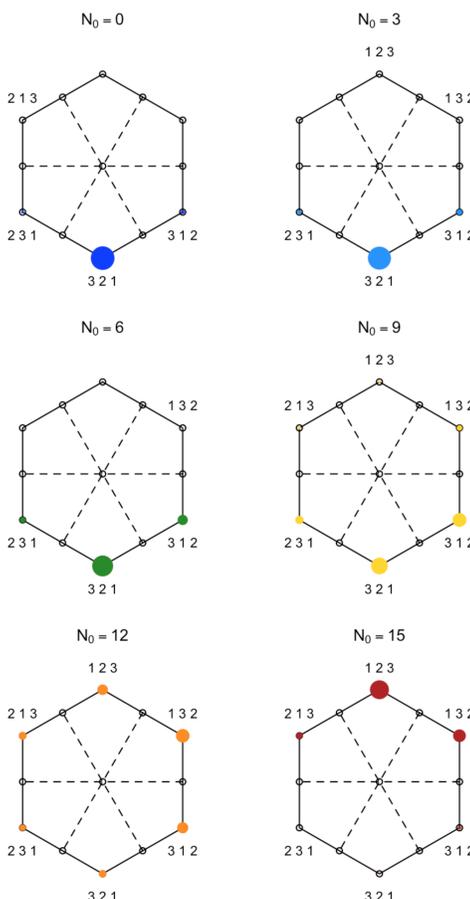


Figure 2: The balls have radius proportional to the frequency of rankings in the posterior sample.

3.A.bis Toy example 2

Same model. Increasing sample size (N). Conjugate prior with $\rho_0 = (1, 2, 3)$ and $\theta N_0 = 3$ (i.e. $N_0 \approx 16$).

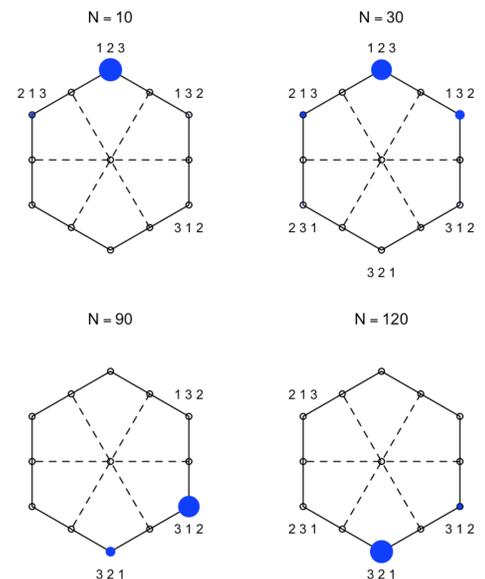


Figure 3: The balls have radius proportional to the frequency of rankings in the posterior sample.

3.B Prior when θ not given

When θ is unknown, the partition function of $\pi(\cdot | \rho_0, \theta N_0)$ depends on the model parameters and cannot be avoided.

Solution: Let $\pi(\theta) \propto Z^*(\theta N_0, \rho_0)$, so that the posterior full conditional is treatable.

3.B Prior when θ not given

Same model. Increasing sample size (N). $\rho_0 = (1, 2, 3)$ and $N_0 = 16$.

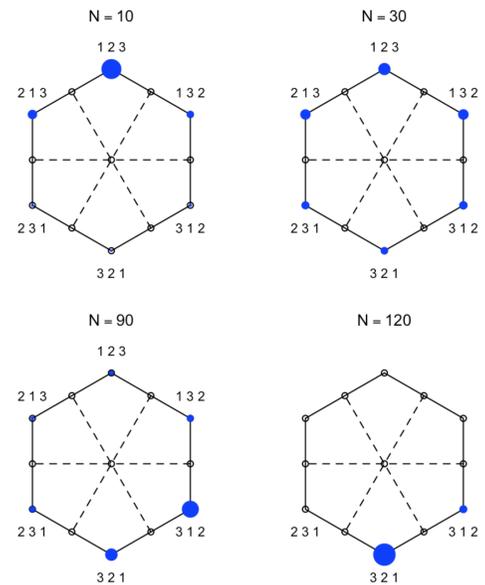


Figure 4: The balls have radius proportional to the frequency of rankings in the posterior sample.

Ongoing work

- Can we say something about the convergence rate? \rightarrow Can we say something about $Z^*(\cdot, \cdot)$?
- Applications?

Interesting data to test our methods on? Please take contact!

References

- Diaconis, P. (1988), *Group representations in probability and statistics*, Vol. 11 of *Lecture Notes - Monograph Series*, Institute of Mathematical Statistics, Hayward, CA, USA.
- Diaconis, P., Ylvisaker, D. et al. (1979), 'Conjugate priors for exponential families', *The Annals of statistics* **7**(2), 269–281.
- Mallows, C. L. (1957), 'Non-null ranking models. I', *Biometrika* **44**(1/2), 114–130.
- Villa, C. & Walker, S. (2015), 'An objective approach to prior mass functions for discrete parameter spaces', *Journal of the American Statistical Association* **110**(511), 1072–1082.
- Vitelli, V., Sørensen, Ø., Crispino, M., Frigessi, A. & Arjas, E. (2018), 'Probabilistic preference learning with the Mallows rank model', *Journal of Machine Learning Research* **18**(158), 1–49.