Conjugate prior in the Mallows model with Spearman distance

M. Crispino\textsuperscript{1,3} and I. Antoniano-Villalobos\textsuperscript{2}
\textsuperscript{1} Université Grenoble Alpes, Luma, CNRS, LJK, 38000 Grenoble, France. \textsuperscript{2} IREA and DCC, Bocconi University, Milan, Italy. \textsuperscript{3} marta.crispino@inria.fr

Introduction

\begin{itemize}
  \item Ranking and comparing items: crucial for collecting information about preferences in many areas: e.g. marketing, business, politics, genetics.
  \item Mallows model: powerful model for rankings
  \item Question: How to include expert opinions into the analysis? How to elicit a meaningful prior on the consensus ranking of a population?
\end{itemize}

1. The Mallows model with Spearman distance

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

$$P(R(\theta, \rho)) = e^{-n \theta \rho(R)} / Z(\theta)$$

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

$\theta \geq 0$: precision parameter

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

$Z(\theta) = \sum_{(R, \rho) \in \mathcal{P} \times \mathcal{P}_n} e^{-n \theta \rho(R)}$: partition function (independent of $\rho$ because of right-invariant of $d(\cdot, \cdot)$)

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

$$d_2(\rho, \sigma) = \sum_{i=1}^{n} (\rho_i - \sigma_i)^2$$

Consider $n$ items, ranked by $N$ assessors.

Denote by $R_i = \{R_{i1}, R_{i2}, \ldots, R_{in}\} \in \mathcal{P}_n$, the ranking of user $i$, $i = 1, \ldots, N$.

Under the Mallows model with Spearman distance (MDS), given $\theta$, the likelihood can be written as

$$P(R_i | R_i, \theta) = \exp \left( -n \theta \sum_{i=1}^{n} d_2(\rho(R_i), \rho) \right)$$

where $R_i = \frac{1}{N} \sum_{i=1}^{N} R_{ij}$, $i = 1, \ldots, n$, is the sample average of the $i$-th rank.

2. Sufficient statistics and mle

The sufficient statistic for $\rho = (\rho_1, \ldots, \rho_n)$, when $\theta$ is known, is $R = (R_1, \ldots, R_n)$.

Proposition 1. Let $R_1, \ldots, R_n : \mathbb{R} \to \mathcal{P}_n$, and define the vector of sample ranks $R$ as above. Assume $R_i \neq R_j$, for each $i \neq j$, and denote by $Y_i = (Y_{i1}, \ldots, Y_{in}) \in \mathcal{P}_n$, the rank vector of $R_i$, i.e. $Y_{ij} = \sum_{i=1}^{n} \mathbb{1}(R_i \leq R_j)$, $i = 1, \ldots, n$.

Then the unique mle of $\rho$ is

$$\rho_{\text{mle}} = \arg\max_{\rho \in \mathcal{P}_n} \sum_{i=1}^{n} R_i = Y(R).$$

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

Definition 1. The permutohedron of order $n$, $\mathcal{P}_n$, 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 $(1, 2, 3, \ldots, n)$. Equivalently, it is the convex hull of the points $\rho \in \mathcal{P}_n$, the set of $n$-dim permutations.

3. The Bayesian Mallows

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

They assume that $\rho$ 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 $\theta$?

A. $\theta$ given: conjugate prior for $\rho$

B. $\theta$ not given: conjugate conditioned on $\theta$ and clever prior on $\theta$ → MH sampling scheme to approximate the posterior

3.A Conjugate prior for $\theta$ (given $\rho$)

Proposition 2. Keeping $\theta$ fixed, the conjugate prior for $\rho$ in MDS is

$$\pi(\rho | \theta, N_0) = \exp \left( -\theta N_0 \sum_{i=1}^{n} (\rho_i - \rho_0) \right) / Z'(N_0, \rho_0)$$

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

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

$$\rho_N = \frac{N_0}{N_0 + \theta} \sum_{i=1}^{n} \rho_i \in \mathcal{P}_n$$

$$\theta_N = \theta (N_0 + N) / (N_0 + N, \theta)$$

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 $\rho_0$. When $\rho_0 = (\frac{1}{n}, \frac{1}{n}, \ldots, \frac{1}{n})$ the conjugate prior reduces to the uniform, for all $\theta_N$.

3.A.2 Toy example 1

Sample $N = 30$ rankings from the MDS with $\rho = (3, 1, 2)$, and $\theta = 0.18$. Conjugate prior with $\rho_0 = (1, 2, 3)$, and varying $\theta_0 = 0.05$.

3.A.2 Toy example 2

Same model. Increasing sample size ($N$). Conjugate prior with $\rho_0 = (1, 2, 3)$ and $\theta_0 = 3$ (i.e. $N_0 = 16$).

3.B Prior when $\theta$ not given

When $\theta$ is unknown, the partition function of $\pi(\rho | \theta, \mathcal{P}_n)$ depends on the model parameters and cannot be avoided.

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

3.B Prior when $\theta$ not given

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

Ongoing work

- Can we say something about the convergence rate?
- Can we say something about $Z'(\cdot)$?
- Applications?
- Interesting data to test out our methods on? Please take contact!

References


Figure 1: The balls have radii proportional to the frequency of rankings in the posterior sample.

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

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

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