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.archives-ouvertes.fr/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.
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

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

- $\rho \in \mathcal{P}_{n}$: consensus ranking, location parameter
- $\theta \geq 0$: precision parameter
- $d(\cdot, \cdot)$: right-invariant (Diaconis 1988) distance
- $Z(\theta) = \sum_{\rho \in \mathcal{P}_{n}} e^{-s(\theta, \rho)}$: partition function (independent of $\rho$ because of right-invariance of $d(\cdot, \cdot)$)

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

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

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

$$P(R_1, \ldots, R_n|\theta, \rho) \propto \exp\left(2N \sum_{i=1}^{n} d(R_i, \rho)\right)$$

where $R_i = \frac{1}{n} \sum_{j=1}^{n} R_{ij}, i = 1, \ldots, n$, 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 $R_1, \ldots, R_n|\theta, \rho \sim \text{Mall}(\theta, \rho)$, and define the vector of sample ranks $R$ as above. Assume $R_i \neq R_j$, for each $i \neq j$, and denote by $Y_i(R) = (Y_{i1}(R), \ldots, Y_{in}(R)) \in \mathcal{P}_{n}$, the rank vector of $R_i$, i.e., $Y_{ij}(R) = \sum_{i \neq j} 1(\leq i)$, $i = 1, \ldots, n$.

Then the unique mle of $\rho$ is

$$\rho_{\text{mle}} = \underset{\rho \in \mathcal{P}_{n}}{\text{argmax}} \frac{1}{n} \sum_{i=1}^{n} R_i = Y(R).$$

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

Definition 1. The permutahedron 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.

Figure 1: Left: The permutahedron of order 3 is a regular hexagon, filling a crosssection of a $2 \times 2 \times 2$ cube. Right: The permutahedron of order 4 is a truncated 16-cell.

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 $\theta$ & clever prior on $\theta$ → MH sampling scheme to approximate the posterior.

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

Proposition 2. Keeping $\theta$ fixed, the conjugate prior for $\rho \in \mathcal{P}_{n}$ is

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

where $\rho_0 \in \mathcal{P}_{n}$, and $N_0 \in \mathbb{N}$. We call $\pi(\rho|\theta, N_0)$ the extended 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):

$$\rho_N = \frac{N_0 N_0 + R}{N_0 + N + N_0} \in \mathcal{P}_{n}$$

$N_0$ 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 $\rho_0 = (\frac{1}{n}, \ldots, \frac{1}{n})$ the conjugate prior reduces to the uniform, for all $\theta$.

3.A.bis 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 $\theta_0 = \theta N_0$.

3.B Prior when $\theta$ not given

When $\theta$ is unknown, the partition function of $\pi(\rho|\theta, N_0)$ 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 $\theta$ not given

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

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

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

References


