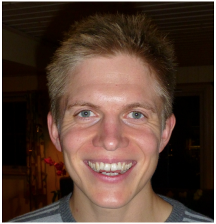
Bayesian preference learning with the Mallows ranking model

Marta Crispino
Mistis team, Inria Grenoble
marta.crispino@inria.fr

Bayesian Statistics in the Big Data Era
CIRM, November, 26th, 2018

Joint work with

Øystein Sørensen
University of Oslo



Valeria Vitelli
University of Oslo

Elja Arjas
University of Helsinki

Arnoldo Frigessi
University of Oslo

Motivation

Preference data is everywhere

- customers express preferences about products and services;
- users select movies on an internet platform (e.g., Netflix);
- genes are ordered based on their expression levels under various experimental conditions.

Motivation

Preference data is everywhere

- customers express preferences about products and services;
- users select movies on an internet platform (e.g., Netflix);
- genes are ordered based on their expression levels under various experimental conditions.

A ranking represents a statement about the relative quality or relevance of the items being ranked: taste, adherence to a specific user profile, relevance to the biological process under investigation.

Motivation

Preference data is everywhere

- customers express preferences about products and services;
- users select movies on an internet platform (e.g., Netflix);
- genes are ordered based on their expression levels under various experimental conditions.

A ranking represents a statement about the relative quality or relevance of the items being ranked: taste, adherence to a specific user profile, relevance to the biological process under investigation.

Assessors rank items: as panels, users, patients.

Ingredients for Ranking data

A set of **items**, to be evaluated. . .

Ingredients for Ranking data

A set of **items**, to be evaluated. . .

. . .and a pool of **assessors** to evaluate them

Ingredients for Ranking data

A set of **items**, to be evaluated. . .

. . .and a pool of **assessors** to evaluate them

A **ranking** is simply a linear ordering of the items

Types of ranking data

FULL RANKINGS

	A1	A2	A3	A4	A5	A6	A7	A8	A9	A10
user 1	7	1	5	6	10	2	4	9	3	8
user 2	5	6	3	4	10	1	8	2	9	7
user 3	3	5	6	2	10	9	7	1	4	8
user 4	4	2	6	9	8	1	7	10	3	5
user 5	5	2	9	1	7	8	4	10	3	6
user 6	3	4	8	5	7	9	2	6	10	1
user 7	5	3	7	4	9	2	6	10	1	8
user 8	1	6	7	9	8	3	4	5	2	10
user 9	2	4	5	10	6	9	3	7	8	1
user 10	9	3	1	7	5	6	4	8	2	10
user 11	3	7	4	5	6	10	2	9	1	8
user 12	2	3	4	8	10	6	9	7	5	1
user 13	3	1	10	7	6	2	8	5	9	4
user 14	3	8	2	5	10	4	7	1	9	6
user 15	5	4	7	2	6	9	3	10	1	8

PARTIAL RANKINGS (top-k)

	A1	A2	A3	A4	A5	A6	A7	A8	A9	A10
user 1	NA	1	NA	NA	NA	2	4	NA	3	NA
user 2	NA	NA	3	4	NA	1	NA	2	NA	NA
user 3	3	NA	NA	2	NA	NA	NA	1	4	NA
user 4	4	2	NA	NA	NA	1	NA	NA	3	NA
user 5	NA	2	NA	1	NA	NA	4	NA	3	NA
user 6	3	4	NA	NA	NA	NA	2	NA	NA	1
user 7	NA	3	NA	4	NA	2	NA	NA	1	NA
user 8	1	NA	NA	NA	NA	3	4	NA	2	NA
user 9	2	4	NA	NA	NA	NA	3	NA	NA	1
user 10	NA	3	1	NA	NA	NA	4	NA	2	NA
user 11	3	NA	4	NA	NA	NA	2	NA	1	NA
user 12	2	3	4	NA	NA	NA	NA	NA	NA	1
user 13	3	1	NA	NA	NA	2	NA	NA	NA	4
user 14	3	NA	2	NA	NA	4	NA	1	NA	NA
user 15	NA	4	NA	2	NA	NA	3	NA	1	NA

PAIRWISE COMPARISONS

user 1	{(A3 < A5), (A7 < A5)}
user 2	{(A2 < A9), (A6 < A5), (A6 < A10), (A8 < A1), (A8 < A7)}
user 3	{(A1 < A9), (A4 < A5), (A4 < A10), (A8 < A7), (A9 < A2)}
user 4	{(A1 < A4), (A2 < A9), (A3 < A4), (A7 < A4), (A9 < A1)}
user 5	{(A4 < A3), (A4 < A7), (A7 < A3), (A7 < A10)}
user 6	{(A2 < A8), (A1 < A2), (A8 < A1)}
user 7	{(A4 < A1), (A9 < A3), (A10 < A5)}
user 8	{(A2 < A4), (A8 < A4), (A9 < A5)}
user 9	{(A1 < A7), (A5 < A9), (A10 < A4), (A10 < A8), (A10 < A9)}
user 10	{(A1 < A10), (A2 < A4), (A3 < A4), (A3 < A5)}
user 11	{(A1 < A8), (A9 < A6)}
user 12	{(A1 < A5), (A7 < A5), (A8 < A7), (A9 < A7), (A10 < A3)}
user 13	{(A2 < A10), (A4 < A7), (A4 < A9), (A6 < A3), (A6 < A5)}
user 14	{(A1 < A4), (A1 < A9)}
user 15	{(A2 < A8), (A3 < A10), (A5 < A6), (A7 < A8), (A9 < A1)}

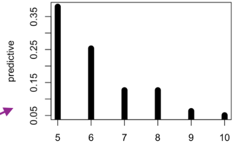
Typical statistical problems

FULL RANKINGS

	A1	A2	A3	A4	A5	A6	A7	A8	A9	A10
user 1	7	1	5	6	10	2	4	9	3	8
user 2	5	6	3	4	10	1	8	2	9	7
user 3	3	5	6	2	10	9	7	1	4	8
user 4	4	2	6	9	8	1	7	10	3	5
user 5	5	2	9	1	7	8	4	10	3	6
user 6	3	4	8	5	7	9	2	6	10	1
user 7	5	3	7	4	9	2	6	10	1	8
user 8	1	6	7	9	8	3	4	5	2	10
user 9	2	4	5	10	6	9	3	7	8	1
user 10	9	3	1	7	5	6	4	8	2	10
user 11	3	7	4	5	6	10	2	9	1	8
user 12	2	3	4	8	10	6	9	7	5	1
user 13	3	1	10	7	6	2	8	5	9	4
user 14	3	8	2	5	10	4	7	1	9	6
user 15	5	4	7	2	6	9	3	10	1	8

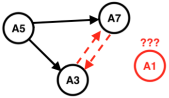
PARTIAL RANKINGS (top-k)

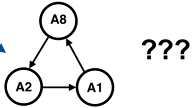
	A1	A2	A3	A4	A5	A6	A7	A8	A9	A10
user 1	NA	1	NA	NA	NA	2	4	NA	3	NA
user 2	NA	NA	3	4	NA	1	NA	2	NA	NA
user 3	3	NA	NA	2	NA	NA	NA	1	4	NA
user 4	4	2	NA	NA	NA	1	NA	NA	3	NA
user 5	NA	2	NA	1	NA	NA	4	NA	3	NA
user 6	3	4	NA	NA	NA	NA	2	NA	NA	1
user 7	NA	3	NA	4	NA	2	NA	NA	1	NA
user 8	1	NA	NA	NA	NA	3	4	NA	2	NA
user 9	2	4	NA	NA	NA	NA	3	NA	NA	1
user 10	NA	3	1	NA	NA	NA	4	NA	2	NA
user 11	3	NA	4	NA	NA	NA	2	NA	1	NA
user 12	2	3	4	NA	NA	NA	NA	NA	NA	1
user 13	3	1	NA	NA	NA	2	NA	NA	NA	4
user 14	3	NA	2	NA	NA	4	NA	1	NA	NA
user 15	NA	4	NA	2	NA	NA	3	NA	1	NA



PAIRWISE COMPARISONS

user 1	{(A3 < A5), (A7 < A5)}
user 2	{(A2 < A9), (A6 < A5), (A6 < A10), (A8 < A1), (A8 < A7)}
user 3	{(A1 < A9), (A4 < A5), (A4 < A10), (A8 < A7), (A9 < A2)}
user 4	{(A1 < A4), (A2 < A9), (A3 < A4), (A7 < A4), (A9 < A1)}
user 5	{(A4 < A3), (A4 < A7), (A7 < A3), (A7 < A10)}
user 6	{(A2 < A8), (A1 < A2), (A8 < A1)}
user 7	{(A4 < A1), (A9 < A3), (A10 < A5)}
user 8	{(A2 < A4), (A8 < A4), (A9 < A5)}
user 9	{(A1 < A7), (A5 < A9), (A10 < A4), (A10 < A8), (A10 < A9)}
user 10	{(A1 < A10), (A2 < A4), (A3 < A4), (A3 < A5)}
user 11	{(A1 < A8), (A9 < A6)}
user 12	{(A1 < A5), (A7 < A5), (A8 < A7), (A9 < A7), (A10 < A3)}
user 13	{(A2 < A10), (A4 < A7), (A4 < A9), (A6 < A3), (A6 < A5)}
user 14	{(A1 < A4), (A1 < A9)}
user 15	{(A2 < A8), (A3 < A10), (A5 < A6), (A7 < A8), (A9 < A1)}





General setting of the Mallows model

- Let $\mathcal{P}_n$, be the space of n-dim permutations
- A **ranking**, $\boldsymbol{R} = (R_1, ..., R_n)$, of n labelled items $\mathcal{A} = \{A_1, ..., A_n\}$ is an element of $\mathcal{P}_n$, where, for all i, R_i is the rank assigned to item A_i.

$$\text{e.g.} \qquad \boldsymbol{R} = \begin{array}{cccccccccc} A_1 & A_2 & A_3 & A_4 & A_5 & A_6 & A_7 & A_8 & A_9 & A_{10} \\ (1, & 7, & 8, & 2, & 10, & 4, & 6, & 9, & 3, & 5\ ) \end{array}$$

General setting of the Mallows model

- Let $\mathcal{P}_n$, be the space of n-dim permutations
- A **ranking**, $\boldsymbol{R} = (R_1, ..., R_n)$, of n labelled items $\mathcal{A} = \{A_1, ..., A_n\}$ is an element of $\mathcal{P}_n$, where, for all i, R_i is the rank assigned to item A_i.

$$\text{e.g.} \quad \boldsymbol{R} = \begin{matrix} A_1 & A_2 & A_3 & A_4 & A_5 & A_6 & A_7 & A_8 & A_9 & A_{10} \\ (1, & 7, & 8, & 2, & 10, & 4, & 6, & 9, & 3, & 5\ ) \end{matrix}$$

- The **Mallows model** (Mallows, 1957) gives the probability density for $\boldsymbol{R} \in \mathcal{P}_n$,

$$P(\boldsymbol{R} \,|\, \alpha, \boldsymbol{\rho}) := \frac{1}{Z_n(\alpha)} \exp\left[-\frac{\alpha}{n} d(\boldsymbol{R}, \boldsymbol{\rho})\right]$$

  - $\boldsymbol{\rho} \in \mathcal{P}_n$: location parameter, shared **consensus ranking**
  - $d(\cdot, \cdot)$: right-invariant (Diaconis, 1988) distance between permutations (example)
  - $\alpha \geq 0$: scale parameter
  - $Z_n(\alpha)$: partition function

General setting of the Mallows model

- Let $\mathcal{P}_n$, be the space of n-dim permutations
- A **ranking**, $\boldsymbol{R} = (R_1, ..., R_n)$, of n labelled items $\mathcal{A} = \{A_1, ..., A_n\}$ is an element of $\mathcal{P}_n$, where, for all i, R_i is the rank assigned to item A_i.

$$\text{e.g.} \quad \boldsymbol{R} = \begin{pmatrix} A_1 & A_2 & A_3 & A_4 & A_5 & A_6 & A_7 & A_8 & A_9 & A_{10} \\ 1, & 7, & 8, & 2, & 10, & 4, & 6, & 9, & 3, & 5 \end{pmatrix}$$

- The **Mallows model** (Mallows, 1957) gives the probability density for $\boldsymbol{R} \in \mathcal{P}_n$,

$$P(\boldsymbol{R} \,|\, \alpha, \boldsymbol{\rho}) := \frac{1}{Z_n(\alpha)} \exp\left[-\frac{\alpha}{n} d(\boldsymbol{R}, \boldsymbol{\rho})\right]$$

  - $\boldsymbol{\rho} \in \mathcal{P}_n$: location parameter, shared **consensus ranking**
  - $d(\cdot, \cdot)$: right-invariant (Diaconis, 1988) distance between permutations (example)
  - $\alpha \geq 0$: scale parameter
  - $Z_n(\alpha)$: partition function

- **Flexibility in the choice of the distance** (driven by the application), (example)
  - **Cayley**, **Hamming**, **Ulam**: measures of disorder $\rightarrow$ genomics, cryptography
  - **Footrule** (l_1), **Spearman** (l_2), **Kendall**: domain of preferences $\rightarrow$ elections, movies

The Mallows model: challenge

$$P(\boldsymbol{R} \mid \alpha, \boldsymbol{\rho}) := \frac{1}{Z_n(\alpha)} \exp\left[-\frac{\alpha}{n} d(\boldsymbol{R}, \boldsymbol{\rho})\right]$$

Challenge for inference: computation of the partition function

$$Z_n(\alpha) = \sum_{\boldsymbol{r} \in \mathcal{P}_n} \exp\left[-\frac{\alpha}{n} d(\boldsymbol{r}, \mathbf{1}_n)\right]$$

$$P(\boldsymbol{R} \mid \alpha, \boldsymbol{\rho}) := \frac{1}{Z_n(\alpha)} \exp\left[-\frac{\alpha}{n} d(\boldsymbol{R}, \boldsymbol{\rho})\right]$$

Challenge for inference: computation of the partition function

$$Z_n(\alpha) = \sum_{\boldsymbol{r} \in \mathcal{P}_n} \exp\left[-\frac{\alpha}{n} d(\boldsymbol{r}, \mathbf{1}_n)\right]$$

With Kendall, Cayley and Hamming distances $\rightarrow Z_n(\alpha)$ has closed form

The Mallows model: challenge

$$P(\boldsymbol{R} \mid \alpha, \boldsymbol{\rho}) := \frac{1}{Z_n(\alpha)} \exp\left[-\frac{\alpha}{n} d(\boldsymbol{R}, \boldsymbol{\rho})\right]$$

Challenge for inference: computation of the partition function

$$Z_n(\alpha) = \sum_{\boldsymbol{r} \in \mathcal{P}_n} \exp\left[-\frac{\alpha}{n} d(\boldsymbol{r}, \mathbf{1}_n)\right]$$

With Kendall, Cayley and Hamming distances $\rightarrow Z_n(\alpha)$ has closed form

With **footrule and Spearman** (i.e. l_1 and l_2) distances $\rightarrow$ no results.
So far solved numerically for very small values of n, as infeasible for larger n.

The Mallows model: challenge

$$P(\boldsymbol{R} \,|\, \alpha, \boldsymbol{\rho}) := \frac{1}{Z_n(\alpha)} \exp\left[-\frac{\alpha}{n} d(\boldsymbol{R}, \boldsymbol{\rho})\right]$$

Challenge for inference: computation of the partition function

$$Z_n(\alpha) = \sum_{\boldsymbol{r} \in \mathcal{P}_n} \exp\left[-\frac{\alpha}{n} d(\boldsymbol{r}, \boldsymbol{1}_n)\right]$$

With Kendall, Cayley and Hamming distances $\rightarrow Z_n(\alpha)$ has closed form

With **footrule and Spearman** (i.e. l_1 and l_2) distances $\rightarrow$ no results.
So far solved numerically for very small values of n, as infeasible for larger n.

Our approach:

1. Strategy to compute $Z_n(\alpha)$ <u>exactly</u> for moderately large values of n.
2. When needed for larger n, Importance Sampling <u>(IS)</u> scheme.

Bayesian inference: full rankings

- N users rank n items $\mathcal{A} = \{A_1, ..., A_n\}$
- Data $\mathbf{R} = \{\mathbf{R}_j\}_{j=1}^N \rightarrow$ full rankings
- $\mathbf{R}_j = (R_{j1}, ..., R_{jn}) \in \mathcal{P}_n$: ranking given by user j to the full set of items
- R_{ji}: rank given to item A_i by user j.

Bayesian inference: full rankings

- N users rank n items $\mathcal{A} = \{A_1, ..., A_n\}$
- Data $\mathbf{R} = \{\mathbf{R}_j\}_{j=1}^N \rightarrow$ full rankings
- $\mathbf{R}_j = (R_{j1}, ..., R_{jn}) \in \mathcal{P}_n$: ranking given by user j to the full set of items
- R_{ji}: rank given to item A_i by user j.
- Statistical model: $\mathbf{R}_1, \ldots, \mathbf{R}_N | \alpha, \boldsymbol{\rho} \overset{i.i.d}{\sim} \text{Mallows}(\alpha, \boldsymbol{\rho})$

$$P(\mathbf{R}_1, \ldots, \mathbf{R}_N; \alpha, \boldsymbol{\rho}) = \frac{1}{Z_n(\alpha)^N} \exp\left\{ -\frac{\alpha}{n} \sum_{j=1}^N d(\mathbf{R}_j, \boldsymbol{\rho}) \right\}$$

- Prior: assume independence between $\boldsymbol{\rho}$ and α and no prior information
  - $\boldsymbol{\rho}$: uniform over $\mathcal{P}_n \rightarrow \pi(\boldsymbol{\rho}) = \frac{1}{n!} 1_{\mathcal{P}_n}(\boldsymbol{\rho})$
  - α: (truncated) exponential prior
- Posterior density

$$\pi(\boldsymbol{\rho}, \alpha | \mathbf{R}_1, \ldots, \mathbf{R}_N) \propto \frac{1}{Z_n(\alpha)^N} \exp\left\{ -\alpha \left[n^{-1} \sum_{j=1}^N d(\mathbf{R}_j, \boldsymbol{\rho}) + \lambda \right] \right\}$$

Bayesian inference: top$-k$ rankings

- N users rank a - **possibly different** - subset of items $\mathcal{A}_j \subseteq \{A_1, A_2, \ldots, A_n\}$
- Typical situation: Each user only assesses her **top**$-k_j$ **preferred items**
- Data $\mathbf{R} = \{\mathbf{R}_j\}_{j=1}^N \rightarrow$ partial rankings

Bayesian inference: top$-k$ rankings

- N users rank a - **possibly different** - subset of items $\mathcal{A}_j \subseteq \{A_1, A_2, \ldots, A_n\}$
- Typical situation: Each user only assesses her **top**$-k_j$ **preferred items**
- Data $\mathbf{R} = \{\mathbf{R}_j\}_{j=1}^N \rightarrow$ partial rankings

Apply **data augmentation** techniques: estimating the lacking ranks consistently with the partial observations.

- Define augmented full rankings $\tilde{\boldsymbol{R}}_1, \ldots, \tilde{\boldsymbol{R}}_N$, where each $\tilde{\boldsymbol{R}}_j$ is compatible with the partial informations in $\boldsymbol{R}_j$

	A_1	A_2	A_3	A_4	A_5
$\mathbf{R}_1$	1	NA	NA	2	3
$\mathbf{R}_2$	1	NA	2	NA	3
$\mathbf{R}_3$	3	1	NA	2	NA
$\mathbf{R}_4$	NA	1	2	3	NA
$\mathbf{R}_5$	NA	1	3	2	NA

$\tilde{\mathbf{R}}_1 = (1, 4, 5, 2, 3)$

$\tilde{\mathbf{R}}_1 = (1, 5, 4, 2, 3)$

?

- Posterior density

$$\pi\left(\alpha, \rho | \boldsymbol{R}_1, \ldots, \boldsymbol{R}_N\right) = \sum_{\tilde{\boldsymbol{R}}_1 \in \mathcal{S}_1} \cdots \sum_{\tilde{\boldsymbol{R}}_N \in \mathcal{S}_N} P\left(\alpha, \rho, \tilde{\boldsymbol{R}}_1, \ldots, \tilde{\boldsymbol{R}}_N | \boldsymbol{R}_1, \ldots, \boldsymbol{R}_N\right).$$

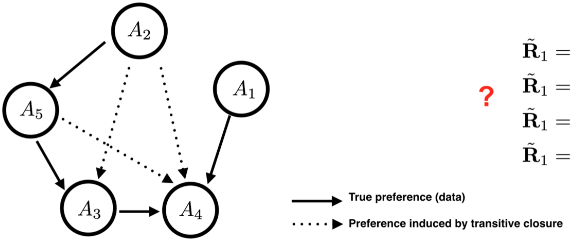
where $\mathcal{S}_j$, set of rankings compatible with $\mathbf{R}_j$, $j = 1, \ldots, N$.

Bayesian inference: transitive pair comparisons

- N users do not see all the possible items, but only express binary preferences between pairs of them
- Data $\{\mathcal{B}_j\}_{j=1}^N$ are sets of pair preferences, of the form $(A_{m_1} \prec A_{m_2})$ if A_{m_1} preferred to A_{m_2}

Bayesian inference: transitive pair comparisons

- N users do not see all the possible items, but only express binary preferences between pairs of them
- Data $\{\mathcal{B}_j\}_{j=1}^N$ are sets of pair preferences, of the form $(A_{m_1} \prec A_{m_2})$ if A_{m_1} preferred to A_{m_2}
- Define augmented full rankings $\tilde{\boldsymbol{R}}_1, \dots, \tilde{\boldsymbol{R}}_N$, where each $\tilde{\boldsymbol{R}}_j$ is compatible with the partial informations in (the transitive closure of) $\mathcal{B}_j$



?

$$\begin{array}{rl} & \;\; A_1 \;\; A_2 \;\; A_3 \;\; A_4 \;\; A_5 \\ \tilde{\mathbf{R}}_1 = & (5,\ 4,\ 2,\ 1,\ 3) \\ \tilde{\mathbf{R}}_1 = & (2,\ 5,\ 3,\ 1,\ 4) \\ \tilde{\mathbf{R}}_1 = & (3,\ 5,\ 2,\ 1,\ 4) \\ \tilde{\mathbf{R}}_1 = & (4,\ 5,\ 2,\ 1,\ 3) \end{array}$$

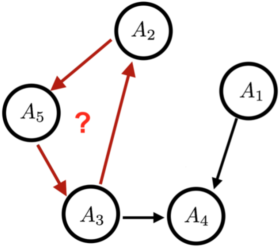
- Posterior density

$$\pi\left(\alpha, \rho | \mathcal{B}_1, \dots, \mathcal{B}_N\right) = \sum_{\tilde{\boldsymbol{R}}_1 \in \text{tc}(\mathcal{B}_1)} \cdots \sum_{\tilde{\boldsymbol{R}}_N \in \text{tc}(\mathcal{B}_N)} P\left(\alpha, \rho | \tilde{\boldsymbol{R}}_1, \dots, \tilde{\boldsymbol{R}}_N\right).$$

Bayesian inference: non-transitive pair comparisons

- Same setting as before BUT users allowed to be inconsistent in their choices
- E.g. It may occur a non-transitive pattern in the data

$$\mathcal{B}_j = \{A_5 \prec A_2,\ A_2 \prec A_3,\ A_3 \prec A_5,\ \ldots\}$$



?

$$\begin{array}{rl} & \quad A_1\ \ A_2\ \ A_3\ \ A_4\ \ A_5 \\ \tilde{\mathbf{R}}_1 = & (5,\ 4,\ 2,\ 1,\ 3) \\ \tilde{\mathbf{R}}_1 = & (2,\ 5,\ 3,\ 1,\ 4) \\ \tilde{\mathbf{R}}_1 = & (3,\ 5,\ 2,\ 1,\ 4) \\ \tilde{\mathbf{R}}_1 = & (4,\ 5,\ 2,\ 1,\ 3) \end{array}$$

... many more ...

- Ideally we want to "coherentize" the preferences, and estimate the latent truth.
- **Idea**: assume non-transitive patterns arise because of mistakes made by the users
- Identification/correction of mistakes: borrowing strength

- Posterior density

$$\pi\left(\alpha, \rho | \mathcal{B}_1, ..., \mathcal{B}_N\right) = \sum_{\tilde{\boldsymbol{R}}_1 \in \mathcal{P}_n} ... \sum_{\tilde{\boldsymbol{R}}_N \in \mathcal{P}_n} P\left(\alpha, \rho | \tilde{\boldsymbol{R}}_1, ..., \tilde{\boldsymbol{R}}_N\right) P\left(\tilde{\boldsymbol{R}}_1, ..., \tilde{\boldsymbol{R}}_N | \mathcal{B}_1, ..., \mathcal{B}_N\right)$$

- Assumption: $P\left(\tilde{\boldsymbol{R}}_1, ..., \tilde{\boldsymbol{R}}_N | \mathcal{B}_1, ..., \mathcal{B}_N\right) = \prod_{j=1}^{N} P\left(\tilde{\boldsymbol{R}}_j | \mathcal{B}_j\right)$
- $P\left(\tilde{\boldsymbol{R}}_j | \mathcal{B}_j\right)$: Weight of each full rank in the sum
- Interpretation: probability of ordering the pairs as in $\mathcal{B}_j$ when the latent ranking for user j is $\tilde{\boldsymbol{R}}_j \rightarrow$ probability of making mistakes in the binary choices
  - **Random mistake**: independent of the pair of items
  - **Logistic model**: the likelihood of a mistake increases if the items are perceived as similar by the user (details)

Implementation and examples

Implementation: Metropolis within Gibbs MCMC, with data augmentation

Many **applications** (require mixture extension):

- **Sushi** benchmark data: full rankings, heterogeneity (*)
- **Meta analysis of gene expression** data: partial rankings (*)
- Preference among **beach pictures**: pairwise comparisons (*)
- **Sound Data**: pairwise comparisons with many non-transitive patterns, due to difficult perception, heterogeneity (*)
- **Movie preferences**: very sparse pairwise comparison data, comparison with Collaborative Filtering (*)

Conclusions

- Ongoing work
  - `R` **package** `BayesMallows`, available on CRAN
  - **Conjugate prior** for ρ (joint work with I. Antoniano-Villalobos) (idea)
  - **Genomics** application: Mixture of Mallows for detection of differential gene expression (joint work with V. Djordjilovic)

Conclusions

- Ongoing work
  - R **package** BayesMallows, available on CRAN
  - **Conjugate prior** for ρ (joint work with I. Antoniano-Villalobos) (idea)
  - **Genomics** application: Mixture of Mallows for detection of differential gene expression (joint work with V. Djordjilovic)
- Future
  - Extension to **rankings with ties** (to model indifference in the preference)
  - Integration of **covariates** (of items and/or of users)
  - Variable selection: rank only the items which are worth being ranked
  - Un-equal quality of assessors

Conclusions

- Ongoing work
  - R **package** BayesMallows, available on CRAN
  - **Conjugate prior** for ρ (joint work with I. Antoniano-Villalobos) (idea)
  - **Genomics** application: Mixture of Mallows for detection of differential gene expression (joint work with V. Djordjilovic)
- Future
  - Extension to **rankings with ties** (to model indifference in the preference)
  - Integration of **covariates** (of items and/or of users)
  - Variable selection: rank only the items which are worth being ranked
  - Un-equal quality of assessors

Thanks for your attention!

Crucial References

Barrett, N., and **Crispino, M.** (2018), 'The Impact of 3-D Sound Spatialisation on Listeners' Understanding of Human Agency in Acousmatic Music', *Journal of New Music Research*, pp 1–17

Crispino, M., Arjas, E., Vitelli, V., Frigessi, A. (2016), 'Recommendation from intransitive pairwise comparisons', *RecSys 2016, Boston, MA, USA, 15th-19th Sept 2016*

Crispino, M., Arjas, E., Barrett, N., Vitelli, V. and Frigessi, A. (2018), 'A Bayesian Mallows approach to non-transitive pair comparison data: how human are sounds?', *Forthcoming in the Annals of Applied Statistics*

Liu*, Q., **Crispino*, M.**, Scheel, I., Vitelli, V. and Frigessi, A. (2019), 'Model-Based Learning from Preference Data', *Annual Review of Statistics and Its Application*, 6(1)

Vitelli, V., Sørensen, Ø., **Crispino, M.**, Frigessi, A. and Arjas, E. (2018), 'Probabilistic preference learning with the Mallows rank model', *Journal of Machine Learning Research, 18(1), pp 5796–5844*.

Other References

Diaconis, P. (1988), 'Group representations in probability and statistics', Vol. 11 of *Lecture Notes - Monograph Series*, Institute of Mathematical Statistics, Hayward, CA, USA.

Mallows, C. L. (1957), 'Non-null ranking models. I', *Biometrika*, **44**(1/2), 114–130.

Mukherjee, S. (2016), 'Estimation in exponential families on permutations', *The Annals of Statistics*, **44**(2), 853–875.

Right-invariance

Definition: Right-invariant distance
A distance function is right-invariant, if $d(\boldsymbol{\rho}_1, \boldsymbol{\rho}_2) = d(\boldsymbol{\rho}_1\boldsymbol{\eta}, \boldsymbol{\rho}_2\boldsymbol{\eta})$ for all $\boldsymbol{\eta}, \boldsymbol{\rho}_1, \boldsymbol{\rho}_2 \in \mathcal{P}_n$, where $\boldsymbol{\rho}\boldsymbol{\eta} = \boldsymbol{\rho} \circ \boldsymbol{\eta} = \boldsymbol{\rho}\boldsymbol{\eta} = (\rho_{\eta_1}, ..., \rho_{\eta_n})$.

Example

- 4 students, (A_1, A_2, A_3, A_4), admitted in a PhD program
- initial ranking $\boldsymbol{\rho}_1 = (1, 3, 4, 2)$ (admission)
- final ranking $\boldsymbol{\rho}_2 = (3, 4, 1, 2)$ (general exam)
- $d(\boldsymbol{\rho_1}, \boldsymbol{\rho_2})$ can be thought of as a measure of the goodness of judgement of the PhD admission board.
- If the students are relabelled in a different ordering, for example (A_4, A_2, A_1, A_3), then $\boldsymbol{\rho}_1\boldsymbol{\eta} = (2, 3, 1, 4)$ and $\boldsymbol{\rho}_2\boldsymbol{\eta} = (2, 4, 3, 1)$, where $\boldsymbol{\eta} = (4, 2, 1, 3)$ determines the relabelling of the students.
- Natural to assume $d(\boldsymbol{\rho_1}, \boldsymbol{\rho_2}) = d(\boldsymbol{\rho_1\eta}, \boldsymbol{\rho_2\eta})$, because the situation depicted is the same.

	A_1	A_2	A_3	A_4
$\boldsymbol{\rho}_1$	1	3	4	2
$\boldsymbol{\rho}_2$	3	4	1	2

$\longrightarrow$

	A_4	A_2	A_1	A_3
$\boldsymbol{\rho}_1\boldsymbol{\eta}$	2	3	1	4
$\boldsymbol{\rho}_2\boldsymbol{\eta}$	2	4	3	1

Consequence of right-invariance
For any $\boldsymbol{\rho}_1, \boldsymbol{\rho}_2 \in \mathcal{P}_n$, it holds $d(\boldsymbol{\rho}_1, \boldsymbol{\rho}_2) = d(\boldsymbol{\rho}_1 \boldsymbol{\rho}_2^{-1}, \mathbf{1}_n)$, where $\mathbf{1}_n = (1, 2, ..., n)$.
Then $Z_n(\alpha, \boldsymbol{\rho})$ is free of $\boldsymbol{\rho}$, as

$$Z_n(\alpha, \boldsymbol{\rho}) = \sum_{\boldsymbol{r} \in \mathcal{P}_n} e^{-\frac{\alpha}{n} d(\boldsymbol{r}, \boldsymbol{\rho})} = \sum_{\boldsymbol{r} \in \mathcal{P}_n} e^{-\frac{\alpha}{n} d(\boldsymbol{r}\boldsymbol{\rho}^{-1}, \mathbf{1}_n)} = \sum_{\boldsymbol{r}' \in \mathcal{P}_n} e^{-\frac{\alpha}{n} d(\boldsymbol{r}', \mathbf{1}_n)} = Z_n(\alpha)$$

Common right-invariant distances between permutations $\boldsymbol{\rho}_1, \boldsymbol{\rho}_2 \in \mathcal{P}_n$

- Footrule (l_1): $d_F(\boldsymbol{\rho}_1, \boldsymbol{\rho}_2) = \sum_{i=1}^n |\rho_{1i} - \rho_{2i}|$
- Spearman (l_2): $d_S(\boldsymbol{\rho}_1, \boldsymbol{\rho}_2) = \sum_{i=1}^n (\rho_{1i} - \rho_{2i})^2$
- Kendall: minimum number of **adjacent transpositions** which convert $\boldsymbol{\rho}_1$ into $\boldsymbol{\rho}_2$
- Cayley: minimum number of **transpositions** which convert $\boldsymbol{\rho}_1$ into $\boldsymbol{\rho}_2$
- Ulam: minimum number of **deletion-insertion** operations to convert $\boldsymbol{\rho}_1$ into $\boldsymbol{\rho}_2$.
- Hamming: minimum number of **substitutions** required to convert $\boldsymbol{\rho}_1$ into $\boldsymbol{\rho}_2$.

Go back

Distances: why important

Consider the following two permutations:

$$\boldsymbol{\sigma} = (1, 2, 3, 4, 5, 6, 7, 8, 9, 10)$$
$$\boldsymbol{\tau} = (9, 10, 3, 4, 5, 6, 7, 8, 1, 2)$$

First and second elements of $\boldsymbol{\sigma}$, are at the bottom of $\boldsymbol{\tau}$.

Distances: why important

Consider the following two permutations:

$$\boldsymbol{\sigma} = (1, 2, 3, 4, 5, 6, 7, 8, 9, 10)$$
$$\boldsymbol{\tau} = (9, 10, 3, 4, 5, 6, 7, 8, 1, 2)$$

First and second elements of $\boldsymbol{\sigma}$, are at the bottom of $\boldsymbol{\tau}$.

If $\boldsymbol{\sigma}$ and $\boldsymbol{\tau}$ represent preferences about movies $\rightarrow$ very different profiles.

Distances: why important

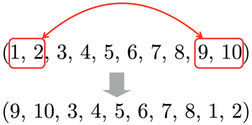
Consider the following two permutations:

$$\boldsymbol{\sigma} = (1, 2, 3, 4, 5, 6, 7, 8, 9, 10)$$
$$\boldsymbol{\tau} = (9, 10, 3, 4, 5, 6, 7, 8, 1, 2)$$

First and second elements of $\boldsymbol{\sigma}$, are at the bottom of $\boldsymbol{\tau}$.

If $\boldsymbol{\sigma}$ and $\boldsymbol{\tau}$ represent preferences about movies $\rightarrow$ very different profiles.

If $\boldsymbol{\sigma}$ and $\boldsymbol{\tau}$ represent genomes $\rightarrow$ just one translocation in the genome



Distances: why important

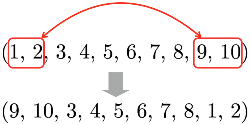
Consider the following two permutations:

$$\boldsymbol{\sigma} = (1, 2, 3, 4, 5, 6, 7, 8, 9, 10)$$
$$\boldsymbol{\tau} = (9, 10, 3, 4, 5, 6, 7, 8, 1, 2)$$

First and second elements of $\boldsymbol{\sigma}$, are at the bottom of $\boldsymbol{\tau}$.

If $\boldsymbol{\sigma}$ and $\boldsymbol{\tau}$ represent preferences about movies $\rightarrow$ very different profiles.

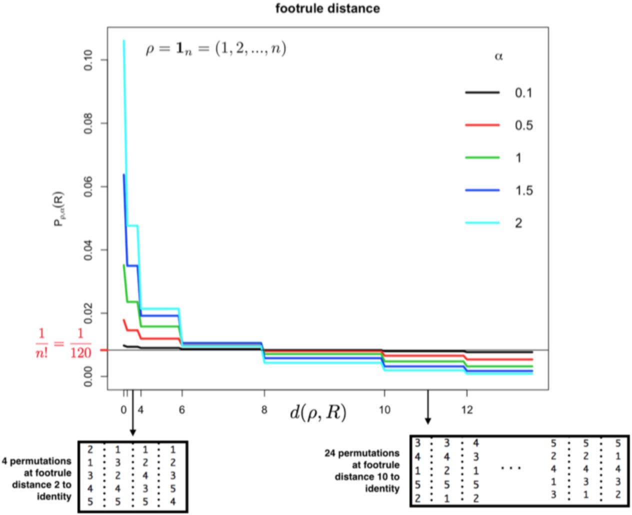
If $\boldsymbol{\sigma}$ and $\boldsymbol{\tau}$ represent genomes $\rightarrow$ just one translocation in the genome



Normalized Spearman (l_2): $d_S(\boldsymbol{\sigma}, \boldsymbol{\tau}) \approx 0.5$
Normalized Cayley: $d_C(\boldsymbol{\sigma}, \boldsymbol{\tau}) \approx 0.28$

Go back

The Mallows density

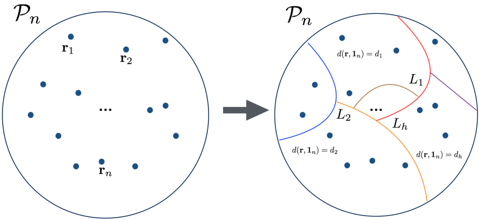


‣ Go back

Exact computation of $Z_n(\alpha)$

$$Z_n(\alpha) = \sum_{\boldsymbol{r} \in \mathcal{P}_n} e^{-\frac{\alpha}{n} d(\boldsymbol{r}, \boldsymbol{1}_n)} = $$

$$= \sum_{d_i \in \mathcal{D}} |L_i| e^{-\frac{\alpha}{n} d_i}$$



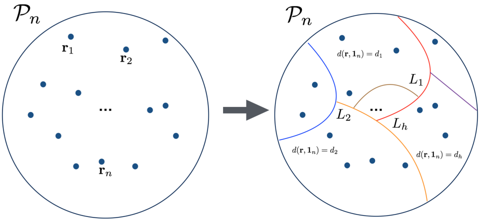
where

- $d(\mathbf{r}, \mathbf{1}_n) \in \mathcal{D} = \{d_1, ..., d_h\}$, h depends on n and $d(\cdot, \cdot)$
- $L_i = \{\mathbf{r} \in \mathcal{P}_n : d(\mathbf{r}, \mathbf{1}_n) = d_i\} \subset \mathcal{P}_n$, $i = 1, ..., h$.

Exact computation of $Z_n(\alpha)$

$$Z_n(\alpha) = \sum_{\mathbf{r} \in \mathcal{P}_n} e^{-\frac{\alpha}{n} d(\mathbf{r}, \mathbf{1}_n)} = $$

$$= \sum_{d_i \in \mathcal{D}} |L_i| e^{-\frac{\alpha}{n} d_i}$$



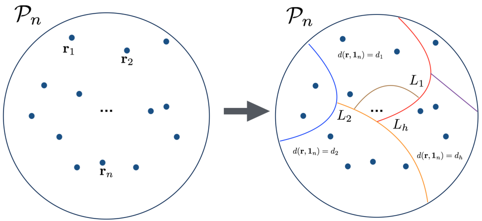
where

- $d(\mathbf{r}, \mathbf{1}_n) \in \mathcal{D} = \{d_1, ..., d_h\}$, h depends on n and $d(\cdot, \cdot)$
- $L_i = \{\mathbf{r} \in \mathcal{P}_n : d(\mathbf{r}, \mathbf{1}_n) = d_i\} \subset \mathcal{P}_n$, $i = 1, ..., h$.

Sufficient to know $|L_i|$, for all values $d_i \in \mathcal{D} \rightarrow$ Easier, but still unfeasible for large n

Exact computation of $Z_n(\alpha)$

$$Z_n(\alpha) = \sum_{\mathbf{r}\in\mathcal{P}_n} e^{-\frac{\alpha}{n} d(\mathbf{r},\mathbf{1}_n)} = $$
$$= \sum_{d_i \in \mathcal{D}} |L_i| e^{-\frac{\alpha}{n} d_i}$$



where

- $d(\mathbf{r},\mathbf{1}_n) \in \mathcal{D} = \{d_1, ..., d_h\}$, h depends on n and $d(\cdot,\cdot)$
- $L_i = \{\mathbf{r} \in \mathcal{P}_n : d(\mathbf{r},\mathbf{1}_n) = d_i\} \subset \mathcal{P}_n$, $i = 1, ..., h$.

Sufficient to know $|L_i|$, for all values $d_i \in \mathcal{D}$ → Easier, but still unfeasible for large n

Special cases solution (from the computer programming field)

- Footrule distance; $\mathcal{D} = \{0, 2, 4, ..., \lfloor n^2/2 \rfloor\}$, $|L_i|$ is the sequence A062869 tabulated for $n \leq 50$ in the On-Line Encyclopedia of Integer Sequences (OEIS)
- Spearman's distance: $\mathcal{D}\{0, 2, 4, ..., 2\binom{n+1}{3}\}$, $|L_i|$ is the sequence A175929 tabulated only until $n \leq 14$ in the OEIS

Go back

Importance Sampling approximation of $Z_n(\alpha)$

Let $\mathbf{R}^1, \ldots, \mathbf{R}^K$ sampled from auxiliary distribution $q(\mathbf{R})$, then

$$\hat{Z}_n(\alpha) = K^{-1} \sum_{k=1}^{K} \exp\left[-(\alpha/n) d(\mathbf{R}^k, \mathbf{1}_n)\right] q(\mathbf{R}^k)^{-1}.$$

Pseudo-likelihood approach: Let $\{i_1, \ldots, i_n\}$ be a uniform sample from $\mathcal{P}_n$, giving the order of the pseudo-likelihood factorization. Then

$$P\left(R_{i_n} | \mathbf{1}_n\right) = \frac{\exp\left[-(\alpha/n) d\left(R_{i_n}, i_n\right)\right] \cdot \mathbb{1}_{[1,\ldots,n]}(R_{i_n})}{\sum_{r_n \in \{1,\ldots,n\}} \exp\left[-(\alpha/n) d\left(r_n, i_n\right)\right]},$$

$$P\left(R_{i_{n-1}} | R_{i_n}, \mathbf{1}_n\right) = \frac{\exp\left[-(\alpha/n) d\left(R_{i_{n-1}}, i_{n-1}\right)\right] \cdot \mathbb{1}_{[\{1,\ldots,n\} \setminus \{R_{i_n}\}]}(R_{i_{n-1}})}{\sum_{r_{n-1} \in \{1,\ldots,n\} \setminus \{R_{i_n}\}} \exp\left[-(\alpha/n) d\left(r_{n-1}, i_{n-1}\right)\right]},$$

$$\vdots$$

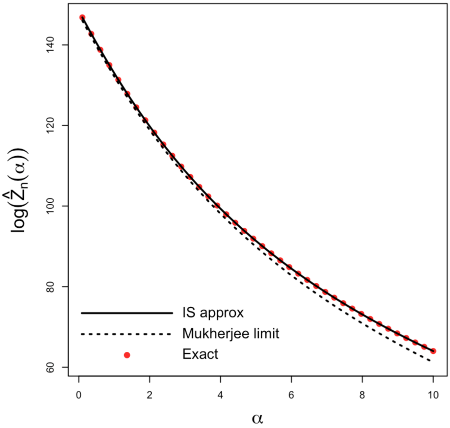
$$P\left(R_{i_2} | R_{i_3}, \ldots, R_{i_n}, \mathbf{1}_n\right) = \frac{\exp\left[-(\alpha/n) d\left(R_{i_2}, i_2\right)\right] \cdot \mathbb{1}_{\left[\{1,\ldots,n\} \setminus \{R_{i_3}, \ldots, R_{i_n}\}\right]}(R_{i_2})}{\sum_{r_2 \in \{1,\ldots,n\} \setminus \{R_{i_3}, \ldots, R_{i_n}\}} \exp\left[-(\alpha/n) d\left(r_2, i_2\right)\right]},$$

$$P\left(R_{i_1} | R_{i_2}, \ldots, R_{i_n}, \mathbf{1}_n\right) = \mathbb{1}_{\left[\{1,\ldots,n\} \setminus \{R_{i_2}, \ldots, R_{i_n}\}\right]}(R_{i_1}).$$

Go back

IS approximation of $Z_n(\alpha)$

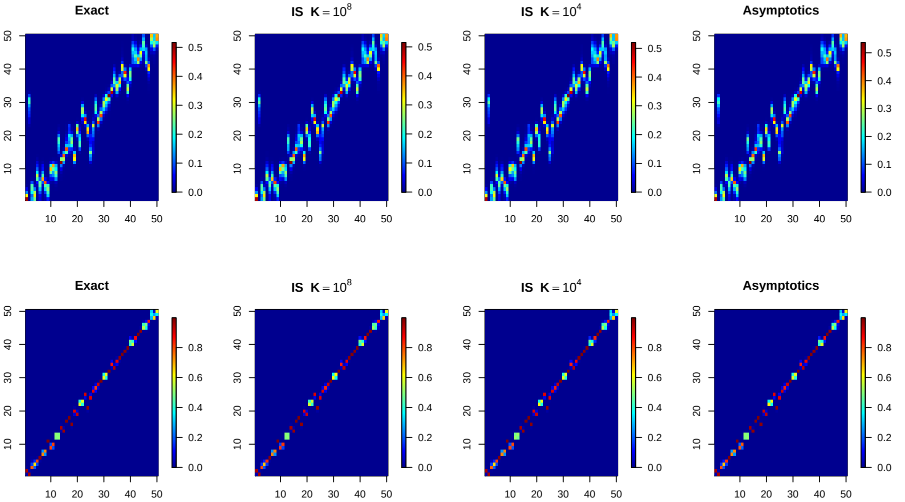
Example: $n = 50$, footrule distance



Mukherjee (2016) limit: asymptotic approximation of $Z_n(\alpha)$

Go back

Effect of the approximation of $Z_n(\alpha)$ on inference



▸ Go back

Conjugate prior for $\boldsymbol{\rho}$ (joint work with I. Antoniano-Villalobos)

Consider a sample of rankings $\boldsymbol{R}_1, ..., \boldsymbol{R}_N | \boldsymbol{\rho}, \theta \overset{i.i.d}{\sim} \mathcal{M}_S(\theta, \boldsymbol{\rho})$, where $\mathcal{M}_S(\cdot, \cdot)$ is the Mallows density with $\theta = \alpha / n$, and Spearman (l_2) distance,

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

Assume θ known, then

$$P(\boldsymbol{R}_1, ..., \boldsymbol{R}_N; \theta, \boldsymbol{\rho}) = \prod_{j=1}^{N} \frac{1}{Z(\theta)} \exp\left\{ -\theta \sum_{i=1}^{n} (R_i - \rho_i)^2 \right\} \propto \exp\left\{ 2\theta N \sum_{i=1}^{n} \rho_i \bar{R}_i \right\},$$

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

Proposition

Let $\mathbb{pp}_n$ be the n-dim permutation polytope, that is, the convex hull of the elements of $\mathcal{P}_n$. Then $\bar{\boldsymbol{R}} = (\bar{R}_1, ..., \bar{R}_n) \in \mathbb{pp}_n$.

Conjugate prior for $\boldsymbol{\rho}$ (joint work with I. Antoniano-Villalobos)

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

$$\pi(\boldsymbol{\rho}|\boldsymbol{\rho}_0, \theta_0) = \frac{1}{Z^*(\theta_0, \boldsymbol{\rho}_0)} \exp\left[-\theta_0 \sum_{i=1}^{n} (\rho_{0i} - \rho_i)^2\right] \mathbb{1}(\boldsymbol{\rho}_0 \in \mathbb{pp}_n) \mathbb{1}(\theta_0 \in \mathbb{R}^+)$$

$$\propto \exp\left[2\theta_0 \sum_{i=1}^{n} \rho_i \rho_{0i}\right]$$

The posterior density for $\boldsymbol{\rho}$ is

$$\pi(\boldsymbol{\rho}|\boldsymbol{R}_1, ..., \boldsymbol{R}_N) \propto \exp\left\{2(\theta_0 + \theta N) \sum_{i=1}^{n} \rho_i \left[\frac{\theta N}{\theta_0 + \theta N} \bar{R}_i + \frac{\theta_0}{\theta_0 + \theta N} \rho_{0,i}\right]\right\}$$

i.e. $\pi(\boldsymbol{\rho}|\boldsymbol{R}_1, ..., \boldsymbol{R}_N)$ same parametric density of the prior, with updated parameters

$$\boldsymbol{\rho}_N = \frac{\theta N}{\theta_0 + \theta N} \bar{\boldsymbol{R}} + \frac{\theta_0}{\theta_0 + \theta N} \boldsymbol{\rho}_0$$

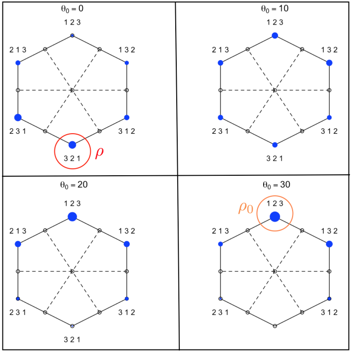
$$\theta_N = \theta_0 + \theta N$$

The result reminds Diaconis and Ylvisaker (1979)

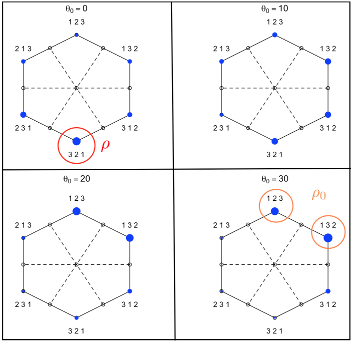
Conjugate prior for ρ (joint work with I. Antoniano-Villalobos)

Example: $n = 3$, $N = 40$, $\theta = 0.5$, $\rho = (3, 2, 1)$.
Sample and obtain $\bar{\boldsymbol{R}} = (2.25, 2.125, 1.625)$.

$\rho_0 = (1, 2, 3)$, varying $\theta_0 = 0, 10, 20, 30$.



$\rho_0 = (1, 2.5, 2.5)$, varying $\theta_0 = 0, 10, 20, 30$.



Go back

Non-transitive pairwise preferences

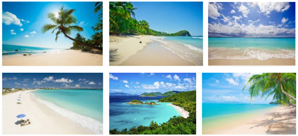
- **Mouse click mistake**:

$$P(\text{mistake} \mid \theta, \boldsymbol{R}_j) = \theta, \quad \theta \in [0, 0.5)$$

- **Logistic model**

$$\text{logit}\, P(\text{mistake} \mid \boldsymbol{R}_j, \beta_0, \beta_1) = -\beta_0 - \beta_1 \frac{d_{\boldsymbol{R}_j, m}}{n-1}$$

where $d_{\boldsymbol{R}_j,m} = |R_{j1} - R_{j2}|$ if $\mathcal{B}_{j,m} = (O_1 \prec O_2)$.



Go back

Sound data: non-transitive pair comparisons (with N. Barrett)

How important is 3-D spatial motion to our understanding of human agency?

- $n = 12$ **abstract sounds**, made from the action of a cellist while playing, each obtained starting at the best representation of the original gesture, and then reducing or removing some aspects of the sound

SOUND1
Full sonification, the best one can make to capture motion - based on what we know about our perception and hearing

SOUND7
Like the previous one, with pitch modulation removed

SOUND10
The 'worst' sonification, spatial variation is flattened, both pitch and volume variations removed.

▸ Go back

Sound data: non-transitive pair comparisons (with N. Barrett)

A group of $N = 46$ **listeners** repeatedly presented with pairs of sounds and asked to choose the one that most evokes the sense of human causation (or physicality)

To what extent listeners report non-transitive sets of preferences?

Sound data: non-transitive pair comparisons (with N. Barrett)

A group of $N = 46$ **listeners** repeatedly presented with pairs of sounds and asked to choose the one that most evokes the sense of human causation (or physicality)

To what extent listeners report non-transitive sets of preferences?
The percentage of listeners who report at least one non-transitivity is **80%**.

Sound data: non-transitive pair comparisons (with N. Barrett)

A group of $N = 46$ **listeners** repeatedly presented with pairs of sounds and asked to choose the one that most evokes the sense of human causation (or physicality)

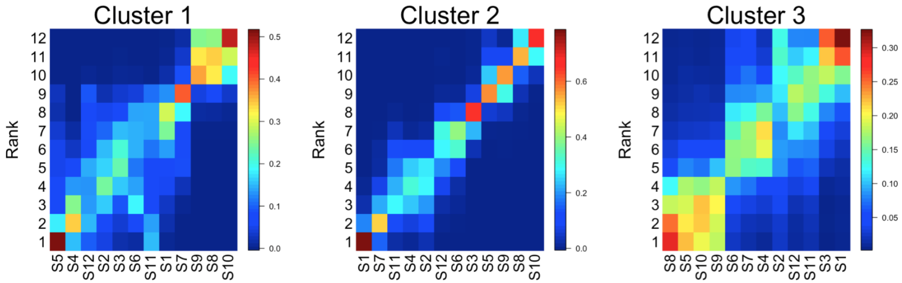
To what extent listeners report non-transitive sets of preferences?
The percentage of listeners who report at least one non-transitivity is **80%**.

We expect the listeners to be **clustered**: differences in the interpretation of the test and in how people listen to sounds → Mixture model generalization of the main model

▸ Go back

Sound data: non-transitive pair comparisons (with N. Barrett)

Posterior consensus ranking ρ of the 3 clusters



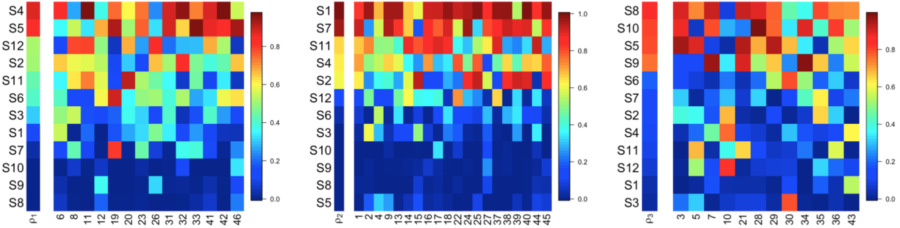
Expert explanation of the clusters:

- Cluster 1: listeners who like slower spatial variation
- Cluster 2: listeners who are listening spatially
- Cluster 3: negative preference for spatial motion

▸ Go back

Sound data: non-transitive pair comparisons (with N. Barrett)

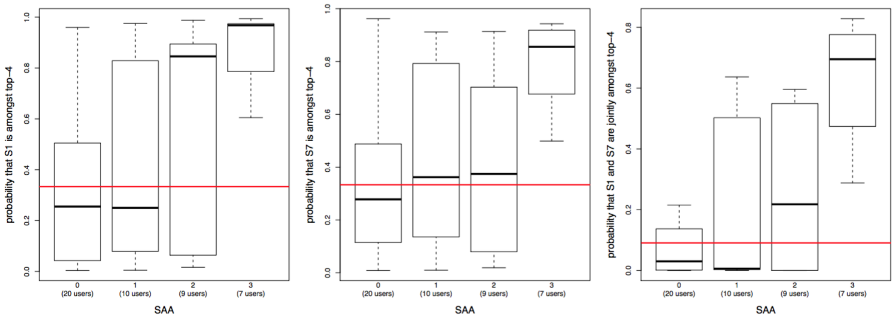
Posterior probabilities for all the sonifications of being ranked among the top-4 for the 3 clusters



▸ Go back

Sound data: non-transitive pair comparisons (with N. Barrett)

Probability that the best sonified sounds are amongst the top-4 ranked sounds (obtained thanks to the estimated individual rankings).
SAA: index measuring listeners' awareness of spatial audio (3 is highly aware)

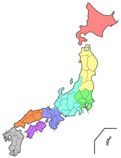


Spatial listening is a skill that is enhanced through training

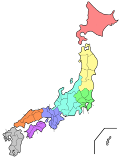
▸ Go back

Sushi data: full rankings

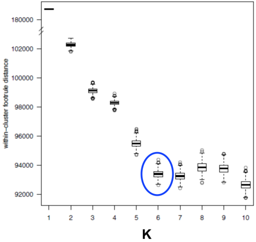
$N = 5000$ Japanese people interviewed: each gives his/her complete ranking of $n = 10$ sushi variants (items)

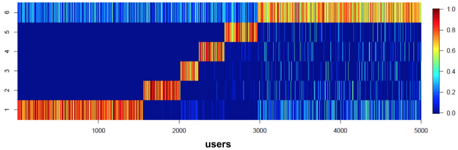


Sushi data: full rankings



$N = 5000$ Japanese people interviewed: each gives his/her complete ranking of $n = 10$ sushi variants (items)



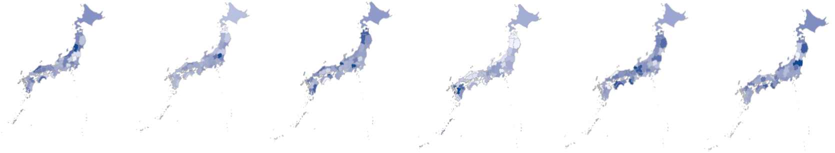


Go back

Sushi data: full rankings

MAP estimate

	$c = 1$	$c = 2$	$c = 3$	$c = 4$	$c = 5$	$c = 6$
τ_c	0.243 (0.23,0.26)	0.131 (0.12,0.14)	0.107 (0.1,0.11)	0.117 (0.11,0.12)	0.121 (0.11,0.13)	0.278 (0.27,0.29)
α_c	3.62 (3.52,3.75)	2.55 (2.35,2.71)	3.8 (3.42,4.06)	4.02 (3.78,4.26)	4.46 (4.25,4.68)	1.86 (1.77,1.94)
1	fatty tuna	shrimp	sea urchin	fatty tuna	fatty tuna	fatty tuna
2	sea urchin	sea eel	fatty tuna	salmon roe	tuna	tuna
3	salmon roe	egg	shrimp	tuna	tuna roll	sea eel
4	sea eel	squid	tuna	tuna roll	shrimp	shrimp
5	tuna	cucumber roll	squid	shrimp	squid	salmon roe
6	shrimp	tuna	tuna roll	egg	sea eel	tuna roll
7	squid	tuna roll	salmon roe	squid	egg	squid
8	tuna roll	fatty tuna	cucumber roll	cucumber roll	cucumber roll	sea urchin
9	egg	salmon roe	egg	sea eel	salmon roe	egg
10	cucumber roll	sea urchin	sea eel	sea urchin	sea urchin	cucumber roll



Go back

Beaches data: pairwise comparisons

- $n = 15$ images of tropical beaches shown in pairs to $N = 60$ users (25 random pairs each)
- Question: "Which of the two beaches would you prefer to go to in your next vacation?"

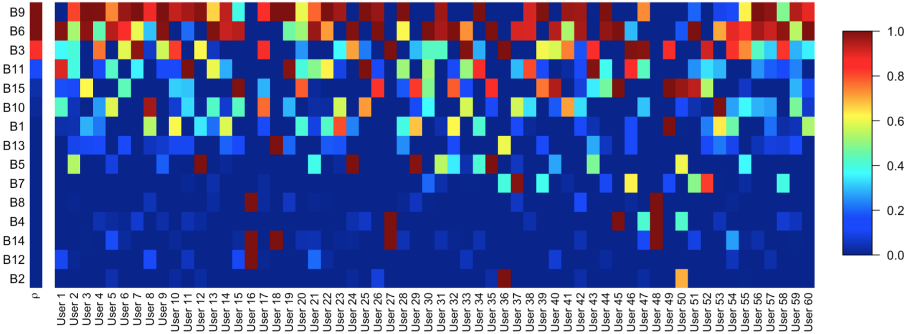
Beaches data: pairwise comparisons

- $n = 15$ images of tropical beaches shown in pairs to $N = 60$ users (25 random pairs each)
- Question: "Which of the two beaches would you prefer to go to in your next vacation?"

Go back

Beaches data: pairwise comparisons

- We can also estimate the individual rankings



Go back

Meta-analysis in Genomics: top$-k$ rankings

Context:

- Studies of differential gene expression between two conditions produce a list of genes, ranked according to their level of differential expression as measured by some test statistics.

Meta-analysis in Genomics: top$-k$ rankings

Context:

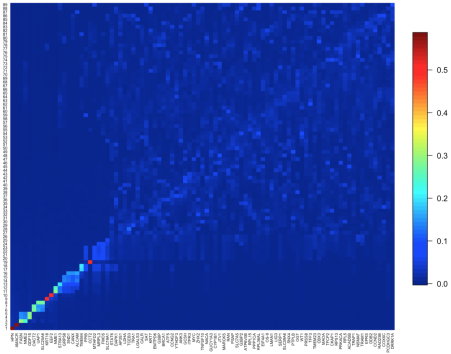
- Studies of differential gene expression between two conditions produce a list of genes, ranked according to their level of differential expression as measured by some test statistics.
- Little agreement among gene lists found by independent studies comparing the same conditions leads to difficulties in finding a consensus list over all available studies. This situation raises the question of whether a consensus top list over all available studies can be found.

Meta-analysis in Genomics: top$-k$ rankings

Context:

- Studies of differential gene expression between two conditions produce a list of genes, ranked according to their level of differential expression as measured by some test statistics.
- Little agreement among gene lists found by independent studies comparing the same conditions leads to difficulties in finding a consensus list over all available studies. This situation raises the question of whether a consensus top list over all available studies can be found.
- Biologists are often concerned with the few most relevant genes in the specific context of the pathology, to set in place further more detailed lab experiments.
- $N = 5$ studies comparing prostate cancer patients with healthy controls, based on differential gene expression
- Each study produces top-25 (i.e. $k = 25$) list of genes (unique genes $n = 89$)

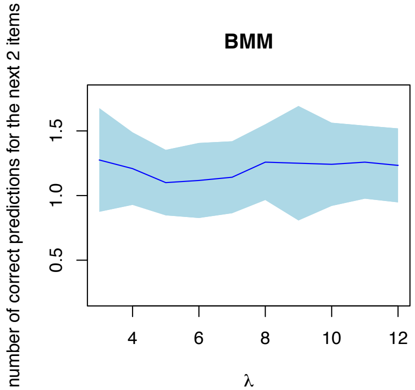
Meta-analysis in Genomics: top$-k$ rankings

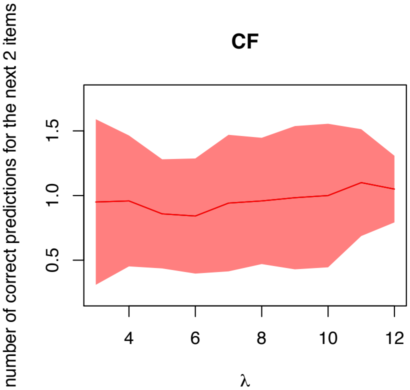


- The fact that $n >> N$, and having partial data, both contribute to keeping precision small
- However, the posterior probability for each gene to be among the top-10 or top-25 is not so low, thus demonstrating that our approach can provide a valid criterion for consensus (with uncertainty quantification).

Go back

Bayesian Mallows model VS Collaborative Filtering (with Q. Liu)





▸ Go back