

A Bayesian Mallows approach to non-transitive pair comparison data: an application to sounds perception

Marta Crispino

Mistis team, Inria Grenoble

`marta.crispino@inria.fr`

CMStatistics

University of Pisa, December, 16th, 2018

Joint work with

Elja Arjas

Dept. of Mathematics,
University of Helsinki



Natasha Barrett

Dept. of Musicology,
University of Oslo



Valeria Vitelli

Dept. of Biostatistics,
University of Oslo



Arnoldo Frigessi

Dept. of Biostatistics,
University of Oslo



Ingredients for preference data

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



Ingredients for preference data

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



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



Ingredients for preference data

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



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



A **ranking** is a linear ordering of the items (representing e.g. individual preferences).



1



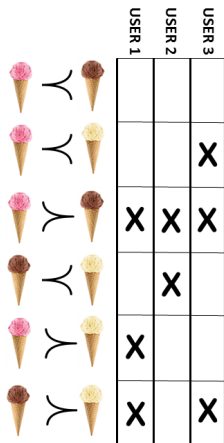
2









3

Basic story

Pair comparisons arise when users state their binary preferences between pairs of items, repeatedly



	USER 1	USER 2	USER 3
			
			X
	X	X	X
		X	
	X		
	X		X

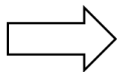
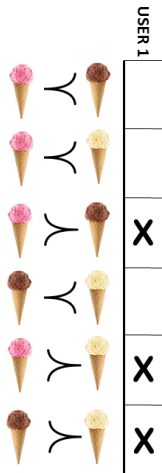
Typical output we collect when

- number of items is large
- items are difficult to rank or rate

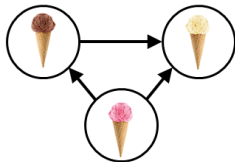
Robust way of collecting data

Basic story

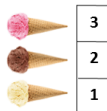
Pair comparisons can lead to a unique ranking.



Items {, , }

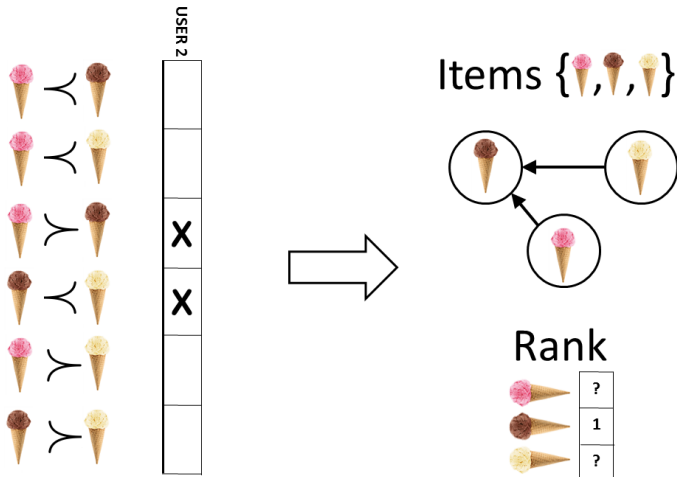


Rank



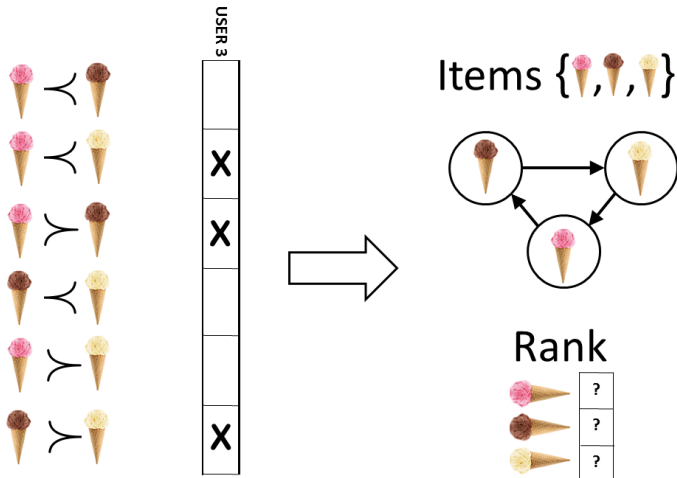
Basic story

Pair comparisons can lead to a partial ranking.



Basic story

Pair comparisons can be **non-transitive**, no ranking available.

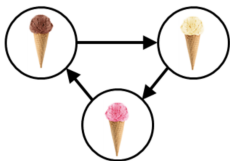


From **non-transitive and incomplete pair comparison** data:

- i) estimate the **modal ranking** of a pool of assessors (rank aggregation task)
- ii) estimate **individual preferences** (personalized recommendation task)
- iii) do i) and ii) when the population is heterogeneous, **cluster** the assessors based on their preferences

Challenge

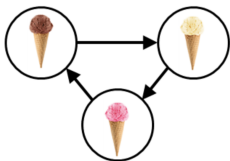
- Ideally we want to “**coherentize**” the preferences, and estimate the latent truth.



- Equivalent to the **minimum feedback arc set problem** on a digraph \equiv remove (or reverse) as few edges as possible so that the remaining graph is acyclic \rightarrow **NP-hard optimization problem**

Challenge

- Ideally we want to “**coherentize**” the preferences, and estimate the latent truth.



- Equivalent to the **minimum feedback arc set problem** on a digraph \equiv remove (or reverse) as few edges as possible so that the remaining graph is acyclic \rightarrow **NP-hard optimization problem**
- Our strategy: sampling rather than optimizing
- Setting: **Bayesian Mallows model for ranking data**, adapted to pair comparisons

General setting of the Mallows model

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

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

General setting of the Mallows model

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

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

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

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

- $\rho \in \mathcal{P}_n$: 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

Bayesian inference: full rankings

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

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

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

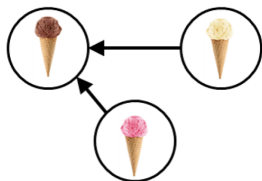
$$\pi(\boldsymbol{\rho}, \alpha | \mathbf{R}_1, \dots, \mathbf{R}_N) \propto \frac{1}{Z_n(\alpha)^N} \exp \left\{ -\alpha \left[n^{-1} \sum_{j=1}^N d(\mathbf{R}_j, \boldsymbol{\rho}) + \lambda \right] \right\}$$

Bayesian inference: 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** $\tilde{\mathbf{R}}_1, \dots, \tilde{\mathbf{R}}_N$, where each $\tilde{\mathbf{R}}_j$ is compatible with the partial informations in the data \mathcal{B}_j



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

?

- Posterior density

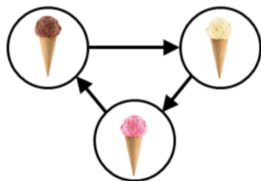
$$\pi(\alpha, \rho | \mathcal{B}_1, \dots, \mathcal{B}_N) = \sum_{\tilde{\mathbf{R}}_1 \in \text{tc}(\mathcal{B}_1)} \dots \sum_{\tilde{\mathbf{R}}_N \in \text{tc}(\mathcal{B}_N)} P(\alpha, \rho | \tilde{\mathbf{R}}_1, \dots, \tilde{\mathbf{R}}_N).$$



When announcing her pairwise preferences, she mentally compares the ranks of the items in her latent ranking \mathbf{R}



Bayesian inference: non-transitive pair comparisons

- Same setting as before BUT users allowed to be inconsistent in their choices

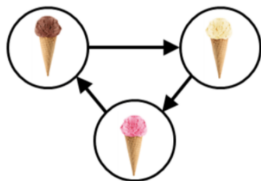




 

$$\begin{aligned} \tilde{\mathbf{R}} &= (1, 2, 3) & \tilde{\mathbf{R}} &= (2, 3, 1) \\ \tilde{\mathbf{R}} &= (1, 3, 2) & \tilde{\mathbf{R}} &= (3, 1, 2) \\ \tilde{\mathbf{R}} &= (2, 1, 3) & \tilde{\mathbf{R}} &= (3, 2, 1) \end{aligned} \quad ?$$

Bayesian inference: non-transitive pair comparisons

- Same setting as before BUT users allowed to be inconsistent in their choices



 $\tilde{\mathbf{R}} = (1, 2, 3)$  $\tilde{\mathbf{R}} = (2, 3, 1)$
 $\tilde{\mathbf{R}} = (1, 3, 2)$ $\tilde{\mathbf{R}} = (3, 1, 2)$?
 $\tilde{\mathbf{R}} = (2, 1, 3)$ $\tilde{\mathbf{R}} = (3, 2, 1)$

- **Idea:** assume non-transitive patterns arise because of **mistakes** made by the users
- Identification/correction of mistakes: borrowing strength

Bayesian inference: non-transitive pair comparisons

- Posterior density

$$\pi(\alpha, \rho | \mathcal{B}_1, \dots, \mathcal{B}_N) = \sum_{\tilde{\mathbf{R}}_1 \in \mathcal{P}_n} \dots \sum_{\tilde{\mathbf{R}}_N \in \mathcal{P}_n} P(\alpha, \rho | \tilde{\mathbf{R}}_1, \dots, \tilde{\mathbf{R}}_N) P(\tilde{\mathbf{R}}_1, \dots, \tilde{\mathbf{R}}_N | \mathcal{B}_1, \dots, \mathcal{B}_N)$$

- Assumption: $P(\tilde{\mathbf{R}}_1, \dots, \tilde{\mathbf{R}}_N | \mathcal{B}_1, \dots, \mathcal{B}_N) = \prod_{j=1}^N P(\tilde{\mathbf{R}}_j | \mathcal{B}_j)$
- $P(\tilde{\mathbf{R}}_j | \mathcal{B}_j)$: 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 $\tilde{\mathbf{R}}_j \rightarrow$ 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: Metropolis within Gibbs MCMC, with data augmentation

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.

Application to sound data

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?

Application to sound data

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%**.

Application to sound data

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)

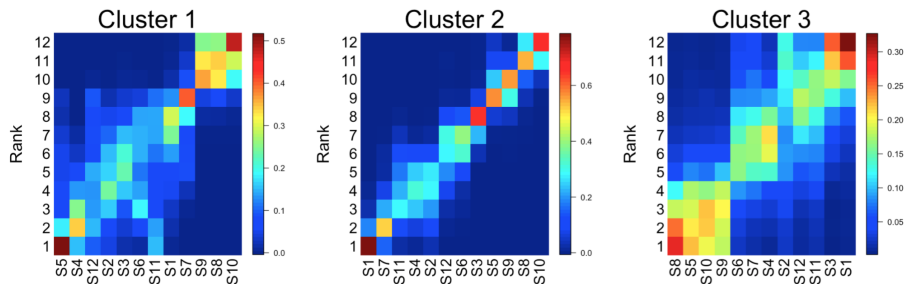
To what extent listeners report non-transitive sets of preferences?

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

Application to sound data

Posterior modal ranking ρ of the 3 clusters

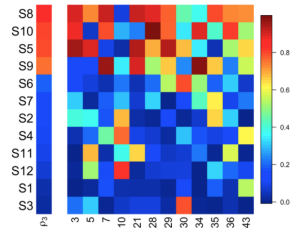
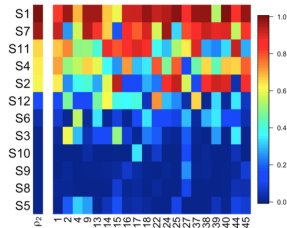
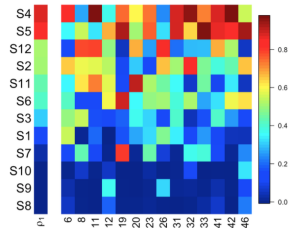


Expert explanation of the clusters:

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

Application to sound data

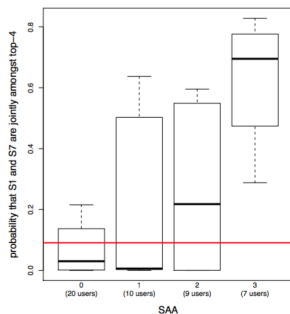
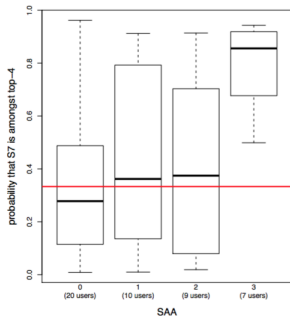
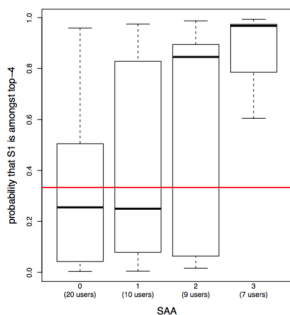
Posterior probabilities for all the sonifications of being ranked among the top-4 for the 3 clusters



Application to sound data

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

Conclusions

- From **non-transitive pair comparisons** of a heterogeneous pool of users: ability to estimate both
 - the modal ranking of each cluster
 - the individual rankings of each user

Conclusions

- From **non-transitive pair comparisons** of a heterogeneous pool of users: ability to estimate both
 - the modal ranking of each cluster
 - the individual rankings of each user
- Ongoing work
 - `R > BayesMallows`, **available on CRAN** (with Ø. Sørensen, V. Vitelli, Q. Liu)
 - **Conjugate prior** for ρ (with I. Antoniano-Villalobos) (idea)

Conclusions

- From **non-transitive pair comparisons** of a heterogeneous pool of users: ability to estimate both
 - the modal ranking of each cluster
 - the individual rankings of each user
- Ongoing work
 - `R > BayesMallows`, **available on CRAN** (with Ø. Sørensen, V. Vitelli, Q. Liu)
 - **Conjugate prior** for ρ (with I. Antoniano-Villalobos) (idea)
- Future
 - Extension to **rankings with ties** (to model indifference in the preference)
 - Integration of **covariates** (of items and/or of users)

Conclusions

- From **non-transitive pair comparisons** of a heterogeneous pool of users: ability to estimate both
 - the modal ranking of each cluster
 - the individual rankings of each user
- Ongoing work
 - `R > BayesMallows`, **available on CRAN** (with Ø. Sørensen, V. Vitelli, Q. Liu)
 - **Conjugate prior** for ρ (with I. Antoniano-Villalobos) ([idea](#))
- Future
 - Extension to **rankings with ties** (to model indifference in the preference)
 - Integration of **covariates** (of items and/or of users)

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 $d(\rho_1, \rho_2) = d(\rho_1\eta, \rho_2\eta)$ for all $\eta, \rho_1, \rho_2 \in \mathcal{P}_n$, where $\rho\eta = \rho \circ \eta = \rho\eta = (\rho_{\eta_1}, \dots, \rho_{\eta_n})$.

Example

- 4 students, (A_1, A_2, A_3, A_4) , admitted in a PhD program
- initial ranking $\rho_1 = (1, 3, 4, 2)$ (admission)
- final ranking $\rho_2 = (3, 4, 1, 2)$ (general exam)
- $d(\rho_1, \rho_2)$ 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 $\rho_1\eta = (2, 3, 1, 4)$ and $\rho_2\eta = (2, 4, 3, 1)$, where $\eta = (4, 2, 1, 3)$ determines the relabelling of the students.
- Natural to assume $d(\rho_1, \rho_2) = d(\rho_1\eta, \rho_2\eta)$, because the situation depicted is the same.

	A_1	A_2	A_3	A_4						
ρ_1	1	3	4	2	\rightarrow	$\rho_1\eta$	2	3	1	4
ρ_2	3	4	1	2		$\rho_2\eta$	2	4	3	1

Consequence of right-invariance

For any $\rho_1, \rho_2 \in \mathcal{P}_n$, it holds $d(\rho_1, \rho_2) = d(\rho_1 \rho_2^{-1}, \mathbf{1}_n)$, where $\mathbf{1}_n = (1, 2, \dots, n)$.
Then $Z_n(\alpha, \rho)$ is free of ρ , as

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

Common right-invariant distances between permutations $\rho_1, \rho_2 \in \mathcal{P}_n$

- Footrule (l_1): $d_F(\rho_1, \rho_2) = \sum_{i=1}^n |\rho_{1i} - \rho_{2i}|$
- Spearman (l_2): $d_S(\rho_1, \rho_2) = \sum_{i=1}^n (\rho_{1i} - \rho_{2i})^2$
- Kendall: minimum number of **adjacent transpositions** which convert ρ_1 into ρ_2
- Cayley: minimum number of **transpositions** which convert ρ_1 into ρ_2
- Ulam: minimum number of **deletion-insertion** operations to convert ρ_1 into ρ_2 .
- Hamming: minimum number of **substitutions** required to convert ρ_1 into ρ_2 .

► Go back

Distances: why important

Consider the following two permutations:

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

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

First and second elements of σ , are at the bottom of τ .

Distances: why important

Consider the following two permutations:

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

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

First and second elements of σ , are at the bottom of τ .

If σ and τ represent preferences about movies \rightarrow very different profiles.

Distances: why important

Consider the following two permutations:

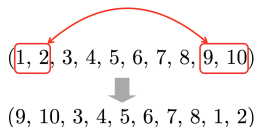
$$\sigma = (1, 2, 3, 4, 5, 6, 7, 8, 9, 10)$$

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

First and second elements of σ , are at the bottom of τ .

If σ and τ represent preferences about movies \rightarrow very different profiles.

If σ and τ represent genomes \rightarrow just one translocation in the genome



Distances: why important

Consider the following two permutations:

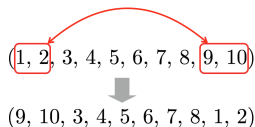
$$\sigma = (1, 2, 3, 4, 5, 6, 7, 8, 9, 10)$$

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

First and second elements of σ , are at the bottom of τ .

If σ and τ represent preferences about movies \rightarrow very different profiles.

If σ and τ represent genomes \rightarrow just one translocation in the genome



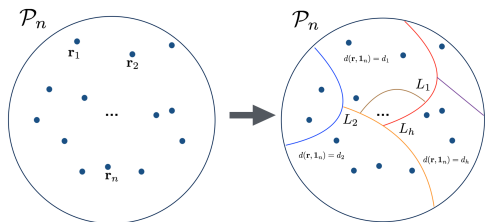
Normalized Spearman (l_2): $d_S(\sigma, \tau) \approx 0.5$

Normalized Cayley: $d_C(\sigma, \tau) \approx 0.28$

► Go back

Exact computation of $Z_n(\alpha)$

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

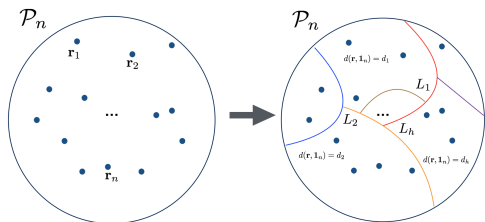


where

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

Exact computation of $Z_n(\alpha)$

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



where

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

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

Importance Sampling approximation of $Z_n(\alpha)$

Let $\mathbf{R}^1, \dots, \mathbf{R}^K$ 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 $\{i_1, \dots, i_n\}$ be a uniform sample from \mathcal{P}_n , giving the order of the pseudo-likelihood factorization. Then

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

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

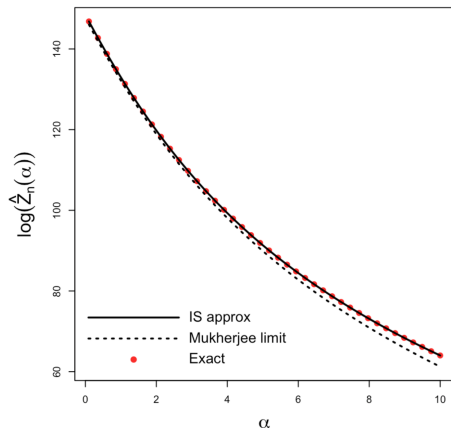
⋮

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

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

IS approximation of $Z_n(\alpha)$

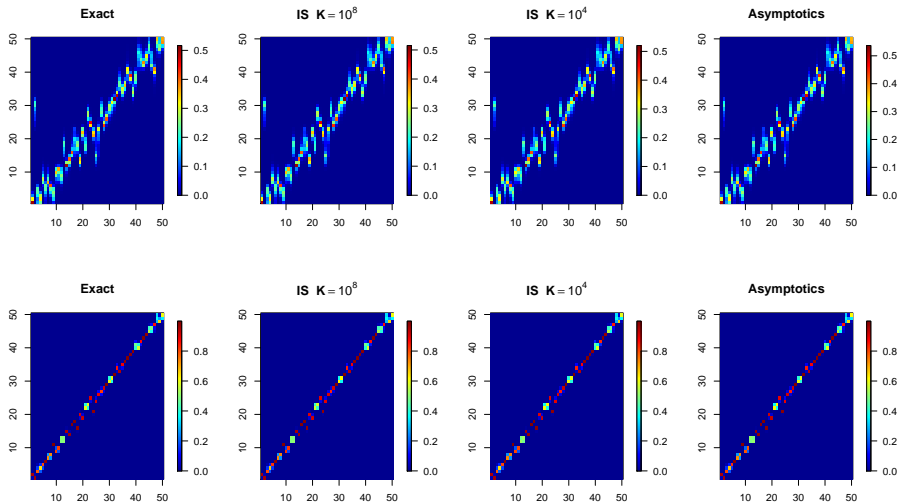
Example: $n = 50$, footrule distance



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

► Go back

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



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

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

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

Assume θ known, then

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

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

Proposition

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

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

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

$$\begin{aligned}\pi(\rho|\rho_0, \theta_0) &= \frac{1}{Z^*(\theta_0, \rho_0)} \exp \left[-\theta_0 \sum_{i=1}^n (\rho_{0i} - \rho_i)^2 \right] \mathbb{1}(\rho_0 \in \mathbb{P}\mathcal{P}_n) \mathbb{1}(\theta_0 \in \mathbb{R}^+) \\ &\propto \exp \left[2\theta_0 \sum_{i=1}^n \rho_i \rho_{0i} \right]\end{aligned}$$

The posterior density for ρ is

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

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

$$\begin{aligned}\rho_N &= \frac{\theta N}{\theta_0 + \theta N} \bar{\mathbf{R}} + \frac{\theta_0}{\theta_0 + \theta N} \rho_0 \\ \theta_N &= \theta_0 + \theta N\end{aligned}$$

The result reminds Diaconis and Ylvisaker (1979)

Non-transitive pairwise preferences

- **Mouse click mistake:**

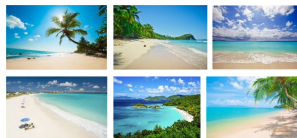
$$P(\text{mistake} \mid \theta, \mathbf{R}_j) = \theta, \quad \theta \in [0, 0.5]$$



- **Logistic model**

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

where $d_{\mathbf{R}_j, m} = |R_{j1} - R_{j2}|$ 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?

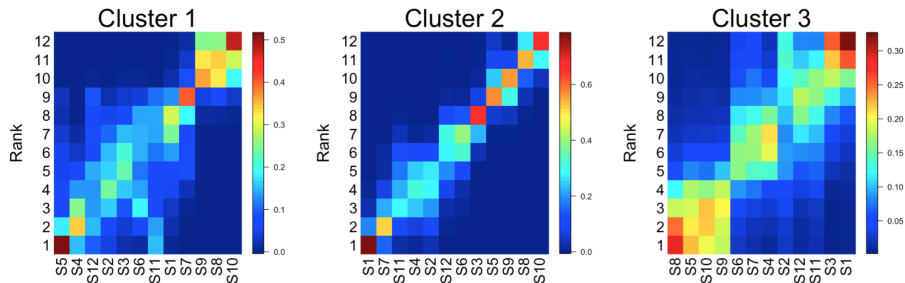
The percentage of listeners who report at least one non-transitivity is **80%**.

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



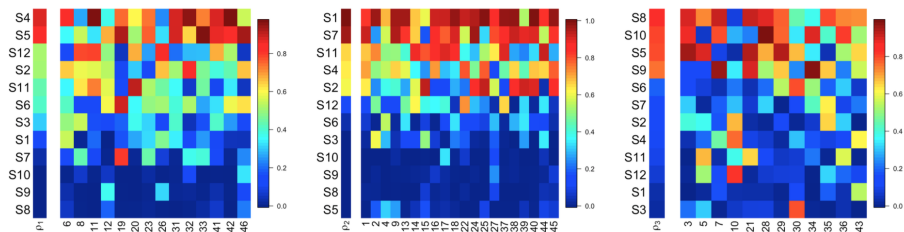
Expert explanation of the clusters:

- 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

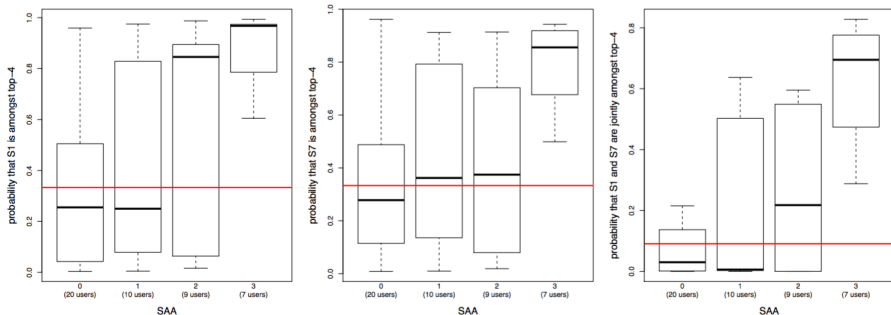


► Go back

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

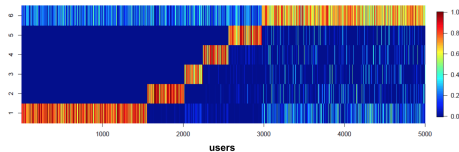
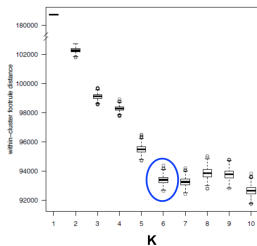
Sushi data: full rankings

$N = 5000$ 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)



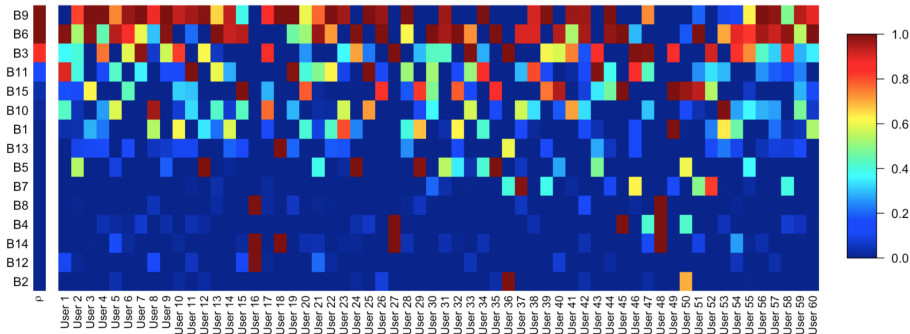
► Go back

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

- We can also estimate the individual rankings



► Go back

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.

Context:

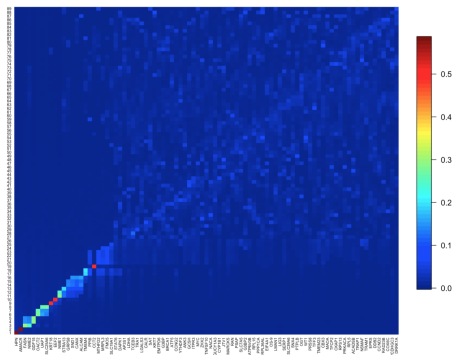
- Studies of differential gene expression between two conditions produce a list of genes, ranked according to their level of differential expression as measured by some test statistics.
- Little agreement among gene lists found by independent studies comparing the same conditions leads to difficulties 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 $n = 89$)

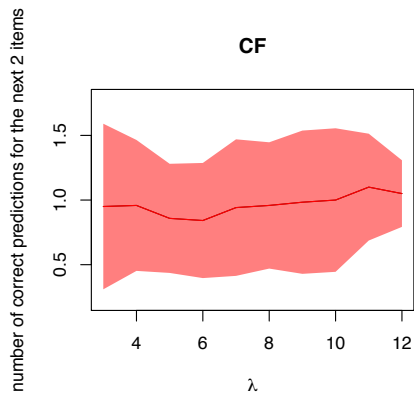
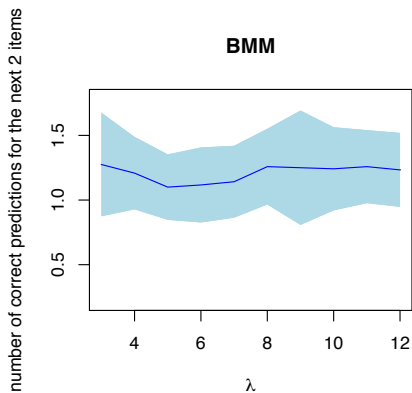
Meta-analysis in Genomics: top- k rankings



- The fact that $n \gg N$, 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