IDA 2014
Thirteenth International Symposium on Intelligent Data Analysis
Leuven, 30 October 2014

Bayesian inference for rank data

Arnoldo Frigessi

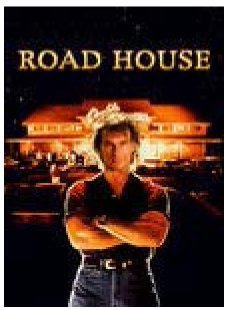
frigessi@medisin.uio.no

OCBE

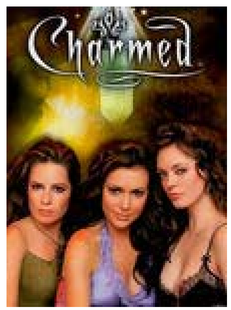
Oslo Centre for Biostatistics and Epidemiology

UiO : University of Oslo

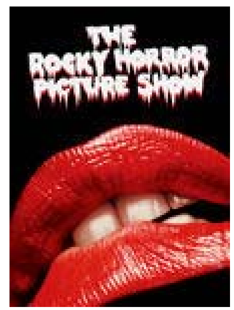
(sfi)^2 Statistics for Innovation


ROAD HOUSE

3


Charmed

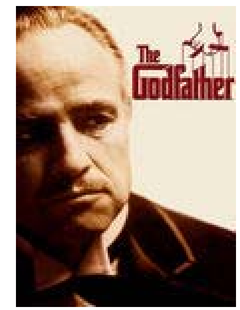
5


THE
ROCKY HORROR
PICTURE SHOW

1


THE ROCK
WALKING
TALL

2


The
Godfather

4

- **Ranked data is everywhere.**

Rankings arise when …

- users express preferences about products and services,
- voters cast ballots in elections,
- research projects are evaluated based on their merits,
- 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.

- **Assessors** rank **items.**
  Designed or observed
  Panel, volunteers, users….

1	3	4	2

Tasks

- Aggregate, merge, summarise multiple rankings to discover shared patterns and structure.

?	?	?	?
1	3	4	2
2	1	3	4
1	2	4	3
4	3	1	2
3	1	4	2
3	1	4	2
1	3	4	2

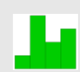
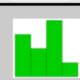
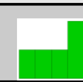
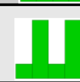
Tasks

- Predict individual ratings, when only partial ratings are made. (not all items rated)

1	3	4	2
2	1	3	4
1	2	4	3
4	3	1	2
?	1	?	2
?	1	?	2
1	?	?	2

Tasks

- Predict individual ratings, when only partial ratings are made. (not all items rated)

1	3	4	2
2	1	3	4
1	2	4	3
4	3	1	2
	1		2
	1		2
1			2

UNCERTAINTY !

Tasks

- Partition assessors in class and predict class membership of new assessors.

1	3	4	2
2	1	3	4
1	2	4	3
4	3	1	2
3	1	4	2
3	1	4	2
1	3	4	2

1	3	4	2
4	3	1	2
3	1	4	2
3	1	4	2

2	1	3	4
1	2	4	3
1	3	4	2

	1		2

movielens

helping you find the *right* movies

- MovieLens is a movie recommendation website.
  You tell us what movies you love and hate.
  We use that information to generate **personalized recommendations** for other movies.

- MovieLens uses *collaborative filtering* to generate recommendations.
  It matches users with similar opinions about movies.
  Each user has a 'neighbourhood' of other like-minded users.
  Ratings from these neighbours are used to create personalized recommendations for the target user.

- Hundreds of thousands of users.
  Started in 1997.
  University of Minnesota.

META ANALYSIS OF GENE EXPRESSIONS ACROSS LABS

- Gene expression is a measure of the activity of a gene in a sample
- ~ 20000 genes measured in a few hundred patients (prostate cancer)
- Repeated in various cohorts and labs with different technologies. Absolute measures cannot be compare. Ranks can.
- Each lab produces a ranked list of genes, relevant for prostate cancer

  Genes = items to be ranked
  Labs = assessors

- Merge the studies to produce a consensus list

Individual top-25 genes from five prostate cancer studies

Rank	Luo(L)	Welsh(W)	Dhana(D)	True(T)	Singh(S)
1	HPN	HPN	OGT	AMACR	HPN
2	AMACR	AMACR	AMACR	HPN	SLC25A6
3	CYP1B1	OACT2	FASN	NME2	EEF2
4	ATF5	GDF15	HPN	CBX3	SAT
5	BRCA1	FASN	UAP1	GDF15	NME2
6	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]

- Mallows model
- Bayesian inference
- MCMC algorithm
- Experiments
    - Full rankings
    - Partial ranking
    - Clustering
    - Pairwise comparisons
    - Preference prediction
- Conclusions

Bayesian inference from rank data

Øystein Sørensen*[1], Valeria Vitelli†[1], Arnoldo Frigessi‡[1,2], and Elja Arjas§[1,3]

arXiv:1405.7945 [stat.ME]

Assume n *items* have been ranked by N *assessors*, such that R_{ij} is the rank given to item i by assessor j. The full data is

$$\mathbf{R}_j = (R_{1j}, R_{2j}, \ldots, R_{nj}), \; j = 1, \ldots, N$$

$\mathcal{P}_n$ be the set of all permutations of $\{1, \ldots, n\}$

$$\mathbf{R}_j \in \mathcal{P}_n$$

Can we find a *consensus ranking*, $\boldsymbol{\rho}$?

$$\boldsymbol{\rho} \in \mathcal{P}_n$$

- Mallows (1957) and Diaconis (1988) proposed the model

$$P(\mathbf{R}|\alpha, \boldsymbol{\rho}) = Z_n(\alpha, \boldsymbol{\rho})^{-1} \exp\{-(\alpha/n) d(\mathbf{R}, \boldsymbol{\rho})\} 1_{\mathcal{P}_n}(\mathbf{R})$$

with distance measure $d(\cdot, \cdot)$ and scale parameter α.

- $1_{\mathcal{P}_n}(\cdot)$ is the indicator function for $\mathcal{P}_n$, the set of all permutations of $\{1, \ldots, n\}$.
- Partition function (more about this later):

$$Z_n(\alpha, \boldsymbol{\rho}) = \sum_{\tilde{\mathbf{R}} \in \mathcal{P}_n} \exp\left\{ \frac{-\alpha}{n} d\left(\tilde{\mathbf{R}}, \boldsymbol{\rho}\right) \right\}.$$

$$P(\mathbf{R}|\alpha, \boldsymbol{\rho}) = Z_n(\alpha, \boldsymbol{\rho})^{-1} \exp\{-(\alpha/n) d(\mathbf{R}, \boldsymbol{\rho})\} 1_{\mathcal{P}_n}(\mathbf{R})$$

- $\alpha = 0$: uniform permutations. `R <- sample(n)`
- $\alpha > 0$: non-uniform permutations, R closer to ρ

$$\boxed{Z_n(\alpha, \boldsymbol{\rho}) = \sum_{\tilde{\mathbf{R}} \in \mathcal{P}_n} \exp\left\{ \frac{-\alpha}{n} d\left(\tilde{\mathbf{R}}, \boldsymbol{\rho}\right) \right\}}$$

If the metric $d(\cdot, \cdot)$ is right-invariant, i.e., invariant to an arbitrary relabeling of the items (Diaconis, 1988, p. 112), the normalizing constant is independent of $\boldsymbol{\rho}$, so we can write

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

where $\mathbf{P}$ denotes an arbitrary permutation in $\mathcal{P}_n$, say $\mathbf{P} = (1, 2, \ldots, n)$. Since this is a sum over $n!$ terms, analytic computation is in general intractable when n is larger than about 10.

- Likelihood for N independent samples

$$P\left(R_1, \ldots, R_N | \alpha, \rho\right) = Z_n\left(\alpha\right)^{-N} \exp\left\{\frac{-\alpha}{n} \sum_{j=1}^{N} d\left(R_j, \rho\right)\right\}$$

- MLE for ρ:

$$\hat{\rho} = \arg\min_{\rho \in \mathcal{P}_n} \left\{\sum_{j=1}^{N} d\left(R_j, \rho\right)\right\}$$

$$\boxed{d(\mathbf{R}, \boldsymbol{\rho})}$$

- The Kendall distance measures the minimum number of pairwise adjacent switches which convert R into ρ.

  Its normalizing constant is given by $Z_n(\alpha) = \prod_{i=1}^{n} \sum_{j=0}^{i-1} e^{-\alpha j/n}$

- Footrule distance: $\|\boldsymbol{\rho} - \mathbf{R}\|_1$
- Spearman distance: $\|\boldsymbol{\rho} - \mathbf{R}\|_2^2$
- and Cayley, Hamming, Ulam, ...

The computation of the normalizing constant in the Mallows model when using other distance measures than Kendall's is NP-complete.

Prior distributions

- when no prior knowledge exists about $\boldsymbol{\rho}$
  uniform distribution over $\mathcal{P}_n$

$$\pi(\boldsymbol{\rho}) = \frac{1}{n!} 1_{\mathcal{P}_n}(\boldsymbol{\rho})$$

- prior distribution for α

$$\pi(\alpha) = \lambda \exp\{-\lambda\alpha\}$$

with hyperparameter $\lambda = 1/10$.

? with hyperparameter $\lambda = 1/10$.

Small chances of very bad assessments:

$P (| R_{ij} - \rho_i | > n/2) = 0.01$

$= \exp (-\alpha/n \; d(R_{ij} , \rho_i))$

(footrule) $= \exp (-\alpha/n \; n/2)$

Which gives $\alpha = 10$.

So let's take exponential with mean 10.

Posterior distribution

$$P(\alpha, \rho | R_1, \ldots, R_N) \propto$$

$$Z_n(\alpha)^{-N} \exp\left[-\alpha\left\{n^{-1}\sum_{j=1}^{N} d(R_j, \rho) + \lambda\right\}\right] 1(\rho \in \mathcal{P}_n)\, 1(\alpha \geq 0)$$

- The posterior summarises all information about the unknown parameters α and ρ
- Point estimation of α and ρ : posterior mean, mode etc.
- Uncertainty about α and ρ : posterior marginals
- Sampling from the posterior by Markov Chain Monte Carlo

- What if j has ranked only top-t? ($R_j \in \mathcal{P}_t, \ \rho \in \mathcal{P}_n$)
- What about confidence statements, e.g., $P(\rho_i \leq 5|\text{data})$?
- What is the precision α?
- What if R_j contains pairwise comparisons rather than rankings?
- Our Bayesian framework handles all of this coherently

Metropolis algorithm

- Symmetric proposal distributions for α and ρ
- Start at $\alpha \geq 0$, $\rho \in \mathcal{P}_n$
- Accept proposal ρ' with probability

$$\min \left\{ 1, \exp \left[-\alpha n^{-1} \sum_{j=1}^{N} \left\{ d\left(R_j, \rho'\right) - d\left(R_j, \rho\right) \right\} \right] 1\left(\rho' \in \mathcal{P}_n\right) \right\}$$

- Accept proposal α' with probability

$$\min \left[1, \frac{Z_n\left(\alpha'\right)^{-N}}{Z_n\left(\alpha\right)^{-N}} \exp \left\{ -\left(\alpha' - \alpha\right) n^{-1} \sum_{j=1}^{N} d\left(R_j, \rho\right) \right\} 1\left(\alpha' \geq 0\right) \right]$$

Symmetric proposal distributions

- Accept proposal ρ' with probability

$$\min\left\{1, \exp\left[-\alpha n^{-1}\sum_{j=1}^{N}\left\{d\left(R_j,\rho'\right)-d\left(R_j,\rho\right)\right\}\right]1\left(\rho'\in\mathcal{P}_n\right)\right\}$$

- ρ' should be close to ρ and in $\mathcal{P}_n$
- We invent the leap-and-shift proposal distribution

Leap-and-shift proposal distribution for ρ

- Choose an item u at random in {1,2,…,n}. Its current rank is ρ_u
- Choose a new rank r for item u, uniformly in ρ_u - L, …, ρ_u + L
  Now two items have rank r and one item (u) has no rank.
- Shift by one all the items of ranks between r and ρ_u

Symmetric proposal distributions

- Accept proposal α' with probability

$$\min\left[1, \frac{Z_n(\alpha')^{-N}}{Z_n(\alpha)^{-N}} \exp\left\{-(\alpha' - \alpha)\, n^{-1} \sum_{j=1}^{N} d(R_j, \rho)\right\} 1(\alpha' \geq 0)\right]$$

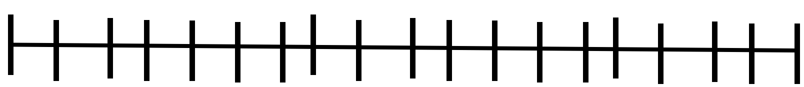
- Simple: $\alpha' \sim N(\alpha, \sigma_\alpha^2)$
- Harder: We need $Z_n(\alpha)$

Partition function

$$Z_n(\alpha) = \sum_{\tilde{\mathbf{R}} \in \mathcal{P}_n} \exp\left\{ \frac{-\alpha}{n} d\left(\tilde{\mathbf{R}}, \boldsymbol{\rho}\right) \right\}$$

ρ=(1,2,3,…, n)

Since $Z_n(\alpha)$ does not depend on $\boldsymbol{\rho}$, the approximation can be done offline for a set of discrete α values, and interpolated to yield an estimate over a continuous range. For each α' proposed in the MCMC algorithm, it is then sufficient to look up the value $\hat{Z}_n(\alpha')$ in our stored estimate.



Importance sampling

$$Z_n(\alpha) = \sum_{R \in \mathcal{P}_n} \frac{\exp\left\{\frac{-\alpha}{n} d(R, \rho)\right\} q(R)}{q(R)}$$

$$= E_q\left\{ \frac{\exp\left\{\frac{-\alpha}{n} d(R, \rho)\right\}}{q(R)} \right\}$$

- For $R^1, \dots, R^K \sim q(R)$,

$$\hat{Z}_n(\alpha) = \frac{1}{K} \sum_{k=1}^{K} \frac{\exp\left\{-\frac{\alpha}{n} d\left(R^k, \rho\right)\right\}}{q\left(R^k\right)}$$

- $\mathbb{E}\left\{\hat{Z}_n(\alpha)\right\} = Z_n(\alpha)$, and if $q(\cdot)$ is close to the target, we get a good estimate with $K << n!$

Importance sampling distribution $q(\cdot)$

- Start with the nth item:

$$P\left(R_n|\rho\right) \propto \exp\left\{\frac{-\alpha}{n} d\left(R_n, \rho_n\right)\right\}$$

- Then the $(n-1)$th item:

$$P\left(R_{n-1}|R_n, \rho\right) \propto \exp\left\{\frac{-\alpha}{n} d\left(R_{n-1}, \rho_{n-1}\right)\right\} 1\left(R_{n-1} \neq R_n\right)$$

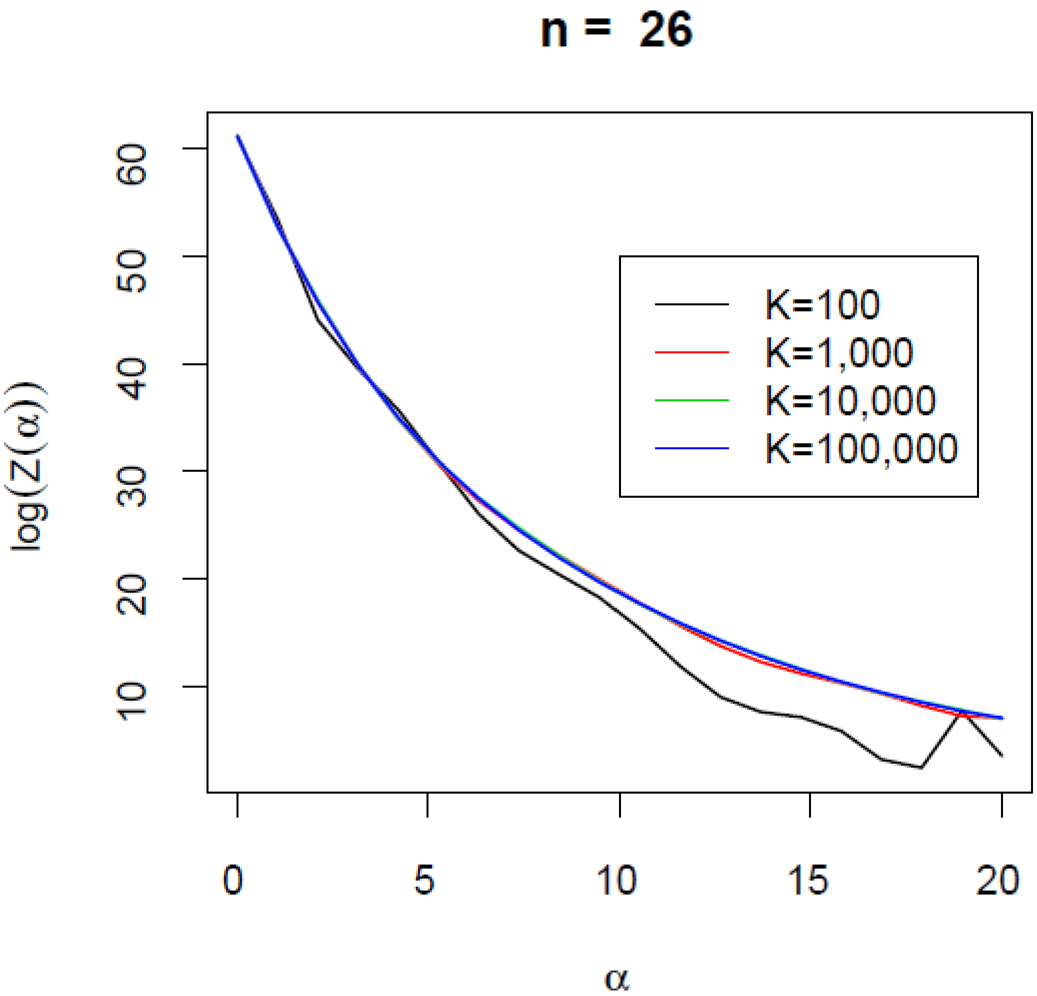
- ... until the 1st item:

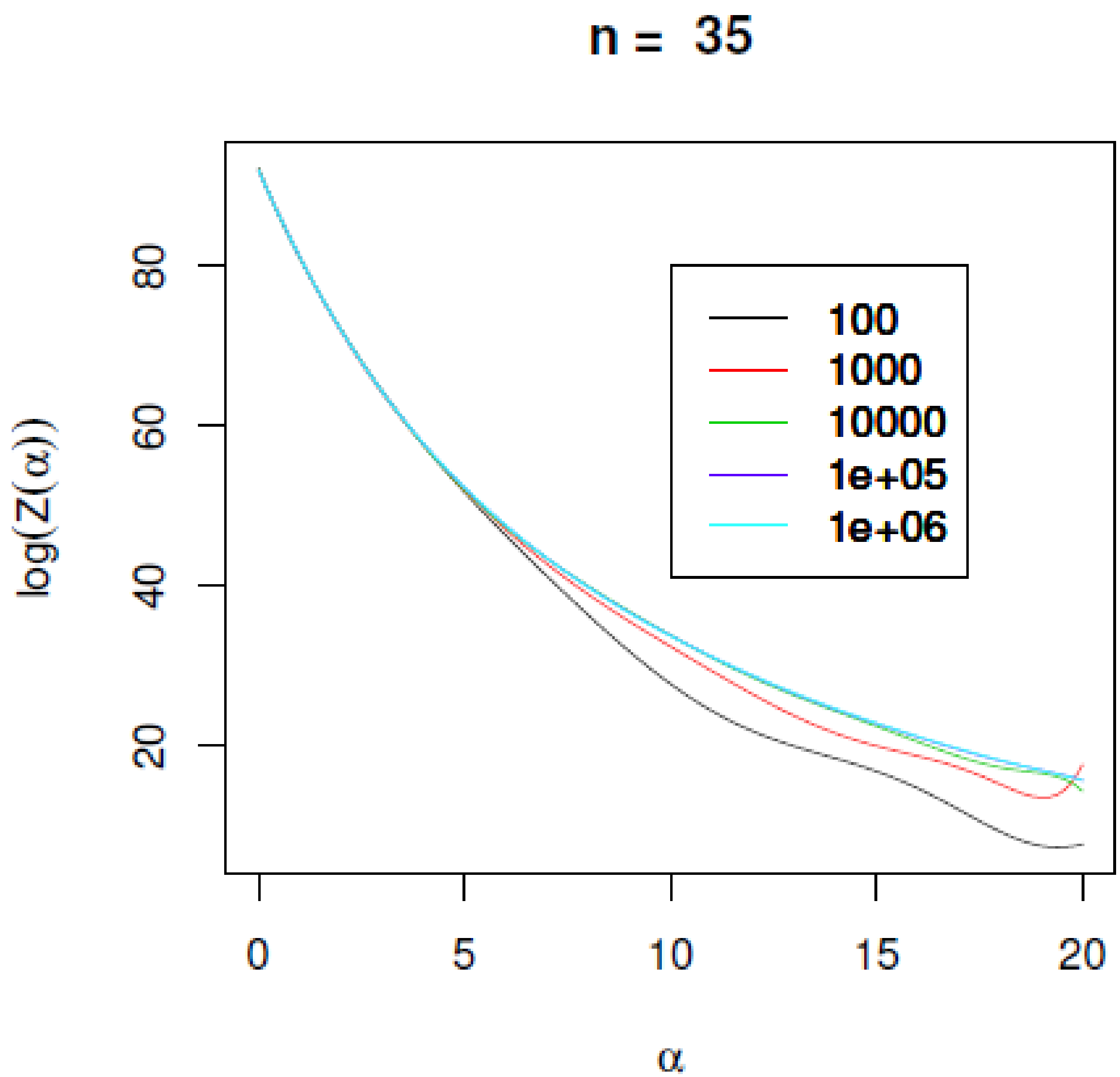
$$R_1 = \{1, \ldots, n\} \setminus \{R_2, \ldots, R_n\}$$

- Now, $R = (R_1, \ldots, R_n) \in \mathcal{P}_n$

(Pseudolikelihood)

- Example with $n = 26$, i.e., $n! > 4 * 10^{26}$




n = 35
100
1000
10000
1e+05
1e+06
log(Z(α))
20
40
60
80
0
5
10
15
20
α

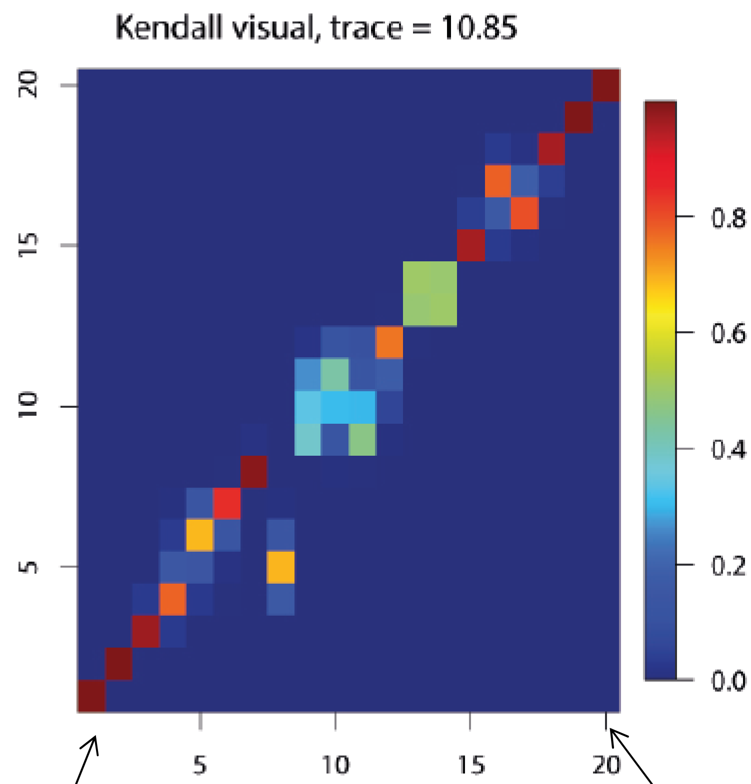
Experiments

The potato experiment

- We bought 20 potatoes, and asked students and colleagues to rank them by weight *without touching*

n=26 students

posterior marginal distribution for the rank of each potato



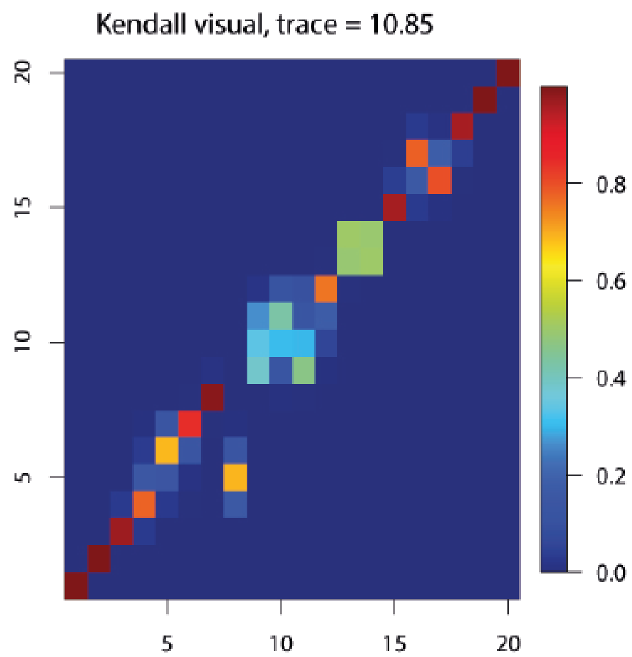
true rank

heaviest potato

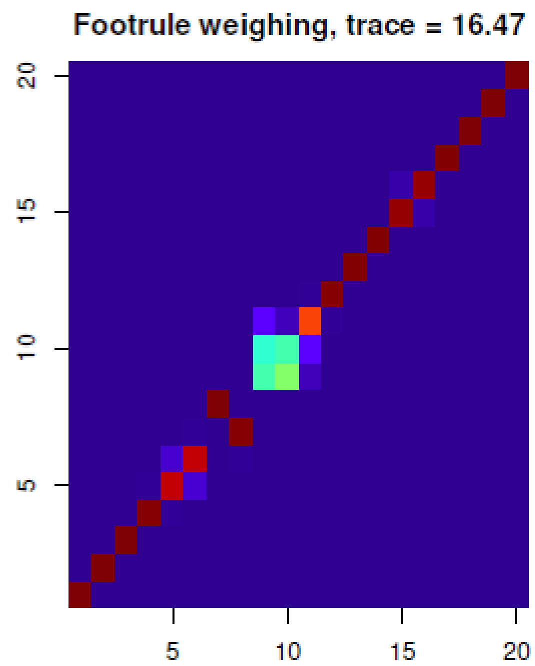
lightest potato

- Represents uncertainty
- Central potatoes are the one ranked with highest uncertainty

by looking



by touching



Less uncertainty

Partially ranked data

- Only a subset of the items have been ranked.
- Ranks can be missing at random, or the assessors may only have ranked, say, the top-5 items.
- Can be handled easily in the Bayesian framework, by applying data augmentation techniques (Tanner and Wong, 1987): estimating the lacking ranks consistently with the partial observations.

$$\tilde{R}_1, \ldots, \tilde{R}_N$$

- Let $R_1, \ldots, R_N$ be the data, with missing values
- MCMC alternates between augmented data,

$$P\left(\tilde{R}_1, \ldots, \tilde{R}_N | \alpha, \rho, R_1, \ldots, R_N\right),$$

  and parameters,

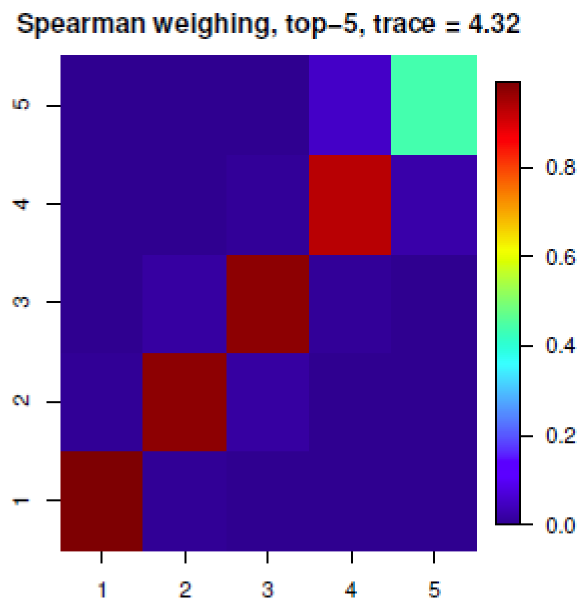
$$P\left(\alpha, \rho | \tilde{R}_1, \ldots, \tilde{R}_N\right).$$

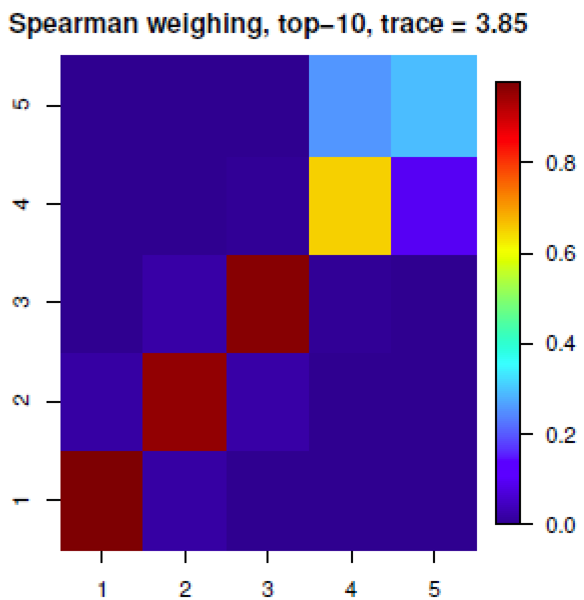
- At convergence, yields samples from

$$P\left(\alpha, \rho | R_1, \ldots, R_N\right)$$

Top-5 estimation in the potato experiment

- We are interested in consensus rank of the top 5 potatoes only.
- Shall we ask assessors to rank the 5 top potatoes?
- Or the top 10 (and then we consider only their top 5)?





- Top 5 is better than top 10
- because in the second case 12 potatoes are involved (instead than 8), and each can enter the top 5

Meta-analysis of gene lists

- Five independent studies of prostate cancer gave five lists of top-25 differentially expressed genes (DeConde et al., 2006; Deng et al., 2014; Lin and Ding, 2009)
- *Cases vs. Controls*

N=5 assessors

n=89 items

Rank	Luo	Welsh	Dhana	True	Singh
1	HPN	HPN	OGT	AMACR	HPN
2	AMACR	AMACR	AMACR	HPN	SLC25A6
3	CYP1B1	OACT2	FASN	NME2	EEF2
4	ATF5	GDF15	HPN	CBX3	SAT
5	BRCA1	FASN	UAP1	GDF15	NME2
6	LGALS3	ANK3	GUCY1A3	MTHFD2	LDHA
7	MYC	KRT18	OACT2	MRPL3	CANX
8	PCDHGC3	UAP1	SLC19A1	SLC25A6	NACA
9	WT1	GRP58	KRT18	NME1	FASN
10	TFF3	PPIB	EEF2	COX6C	SND1
11	MARCKS	KRT7	STRA13	JTV1	KRT18
12	OS−9	NME1	ALCAM	CCNG2	RPL15
13	CCND2	STRA13	GDF15	AP3S1	TNFSF10
14	NME1	DAPK1	NME1	EEF2	SERP1
15	DYRK1A	TMEM4	CALR	RAN	GRP58
16	TRAP1	CANX	SND1	PRKACA	ALCAM
17	FMO5	TRA1	STAT6	RAD23B	GDF15
18	ZHX2	PRSS8	TCEB3	PSAP	TMEM4
19	RPL36AL	ENTPD6	EIF4A1	CCT2	CCT2
20	ITPR3	PPP1CA	LMAN1	G3BP	SLC39A6
21	GCSH	ACADSB	MAOA	EPRS	RPL5
22	DDB2	PTPLB	ATP6V0B	CKAP1	RPS13
23	TFCP2	TMEM23	PPIB	LIG3	MTHFD2
24	TRAM1	MRPL3	FMO5	SNX4	G3BP2
25	YTHDF3	SLC19A1	SLC7A5	NSMAF	UAP1

(89 genes in total)

Rank	Gene	$P(\rho \leq \text{Rank})$	$P(\rho \leq 10)$
1	HPN	0.05	0.35
2	AMACR	0.06	0.27
3	GDF15	0.07	0.23
4	NME1	0.09	0.23
5	FASN	0.1	0.2
6	EEF2	0.12	0.19
7	UAP1	0.13	0.18
8	KRT18	0.14	0.17
9	OACT2	0.14	0.15
10	NME2	0.14	0.14

VERY UNCERTAIN!
WEAK CONSENSUS!
N=5 too small

1. Find the gene with highest posterior probability of having rank 1.
2. Among the remaining genes, find the gene with highest posterior probability of having rank 1 or 2.
3. Etc. → cumulative probability

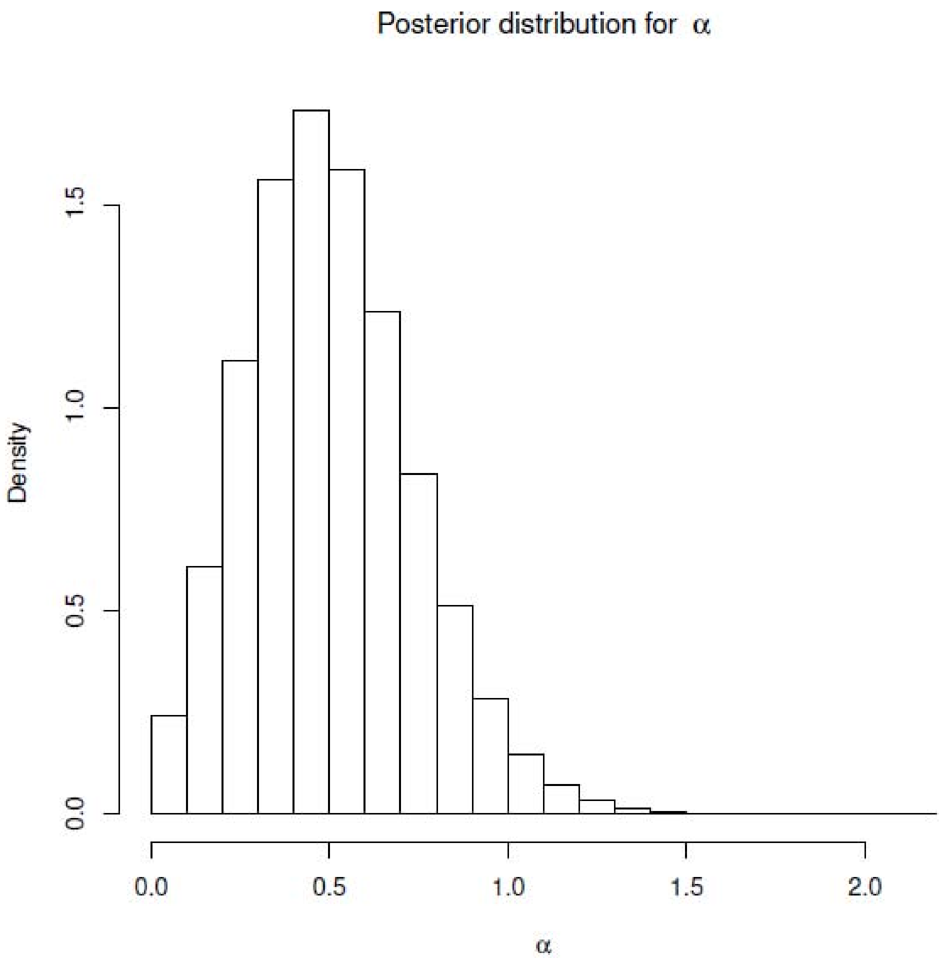
- The probability of being among the top-10 for each gene.

Rank	Gene	$P(\rho \leq \text{Rank})$	$P(\rho \leq 10)$
1	HPN	0.05	0.35
2	AMACR	0.06	0.27
3	GDF15	0.07	0.23
4	NME1	0.09	0.23
5	FASN	0.1	0.2
6	EEF2	0.12	0.19
7	UAP1	0.13	0.18
8	KRT18	0.14	0.17
9	OACT2	0.14	0.15
10	NME2	0.14	0.14

DeConde et al.

Rank	MC4	
1	HPN	0.070
2	AMACR	0.062
3	GDF15	0.041
4	NME1	0.040
5	SLC25A6	0.038
6	KRT18	0.037
7	EEF2	0.037
8	FASN	0.032
9	GUCY1A3	0.031
10	SND1	0.029

- “stationary distribution”, level of consensus
- No precise interpretation.



α is small,

indicating a low level of agreement between the studies.

Clustering the assessors via mixtures

- Assessors not one homogeneous group, but C groups
- We use a mixture of Mallows models to cluster a sample of N assessors according to how they rank the n items.
- We estimate a latent ranking of the items for each cluster of assessors.
- The variables $z_1, \ldots, z_N \in \{1, \ldots, C\}$ assign each assessor to one of the C clusters.
- Prior: Dirichlet distribution on the probabilities that an assessor is in each class
- Each cluster has its own α_c and ρ_c

- Augmented data formulation of the likelihood for the observed rankings $\mathbf{R}_1, \ldots, \mathbf{R}_N$ (assuming conditional independence across clusters):

$$P(\mathbf{R}_1, \ldots, \mathbf{R}_N | \boldsymbol{\rho}_1, \ldots, \boldsymbol{\rho}_C; \alpha_1, \ldots, \alpha_C; z_1, \ldots, z_N) =$$
$$\prod_{j=1}^{N} \frac{1}{Z_n(\alpha_{z_j})} \exp\left(\frac{-\alpha_{z_j}}{n} d(\mathbf{R}_j, \boldsymbol{\rho}_{z_j}) \right) \cdot 1_{\mathcal{P}_n(\mathbf{R}_j)}.$$

- **Location parameters** $\boldsymbol{\rho}_1, \ldots, \boldsymbol{\rho}_C$ (assumed a priori independent) are assigned the same uniform prior.

  Scale parameters $\alpha_1, \ldots, \alpha_C$ follow marginally the exponential prior, with the additional constraint needed for cluster identifiability

$$\pi(\alpha_1, \ldots, \alpha_C) \propto \lambda^C \cdot \exp\left\{ -\lambda \sum_{c=1}^{C} \alpha_c \right\} \cdot 1_{\{\alpha_1 < \ldots < \alpha_C\}}(\alpha_1, \ldots, \alpha_C).$$

- **Cluster labels** are assigned independently according to the prior

$$P(z_1, \ldots, z_N | \tau_1, \ldots, \tau_C) = \prod_{j=1}^{N} \tau_{z_j},$$

(τ_c = probability that an assessor belongs to the c-th cluster).

Cluster probabilities $\tau_1, \ldots, \tau_C$ get the standard symmetric Dirichlet

$$\pi\left(\tau_1, \ldots, \tau_C\right) = \frac{\Gamma(\psi C)}{\Gamma(\psi)^C} \prod_{c=1}^{C} \tau_c^{\psi - 1}.$$

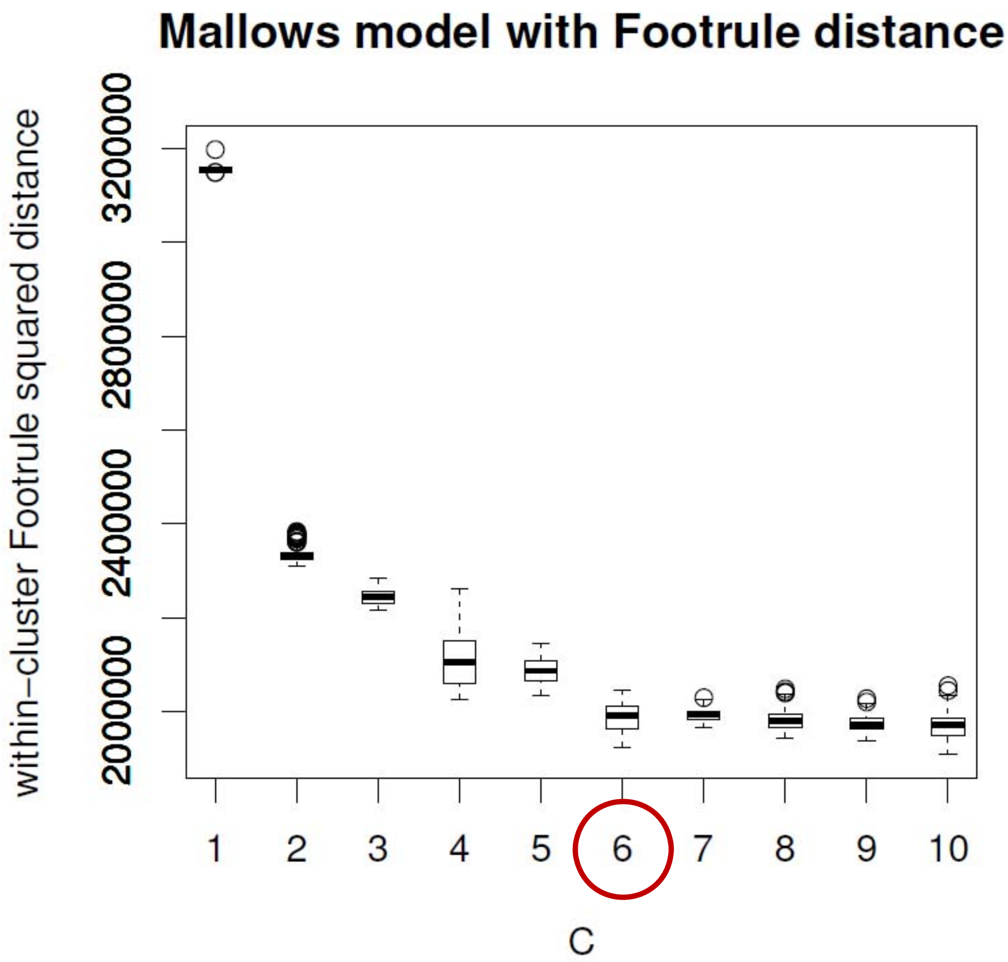
- **posterior distribution**

- N = 5000 people (assessors) were interviewed, each giving his/her complete ranking of n = 10 sushi variants (items):

ebi (shrimp),
anago (sea eel),
maguro (tuna),
ika (squid),
uni (sea urchin),
sake (salmon roe),
tamago (egg),
toro (fatty tuna),
tekka-maki (tuna roll),
kappa-maki (cucumber roll).

within cluster distance
of each rank to the
cluster centroid

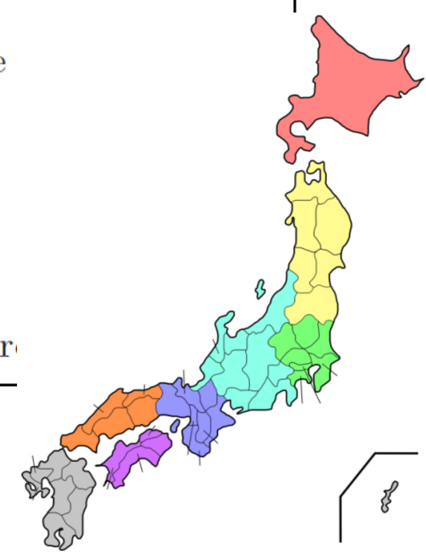
$$\sum_{c=1}^{C} \sum_{j:z_j=c} d(R_j, \rho_c)^2$$



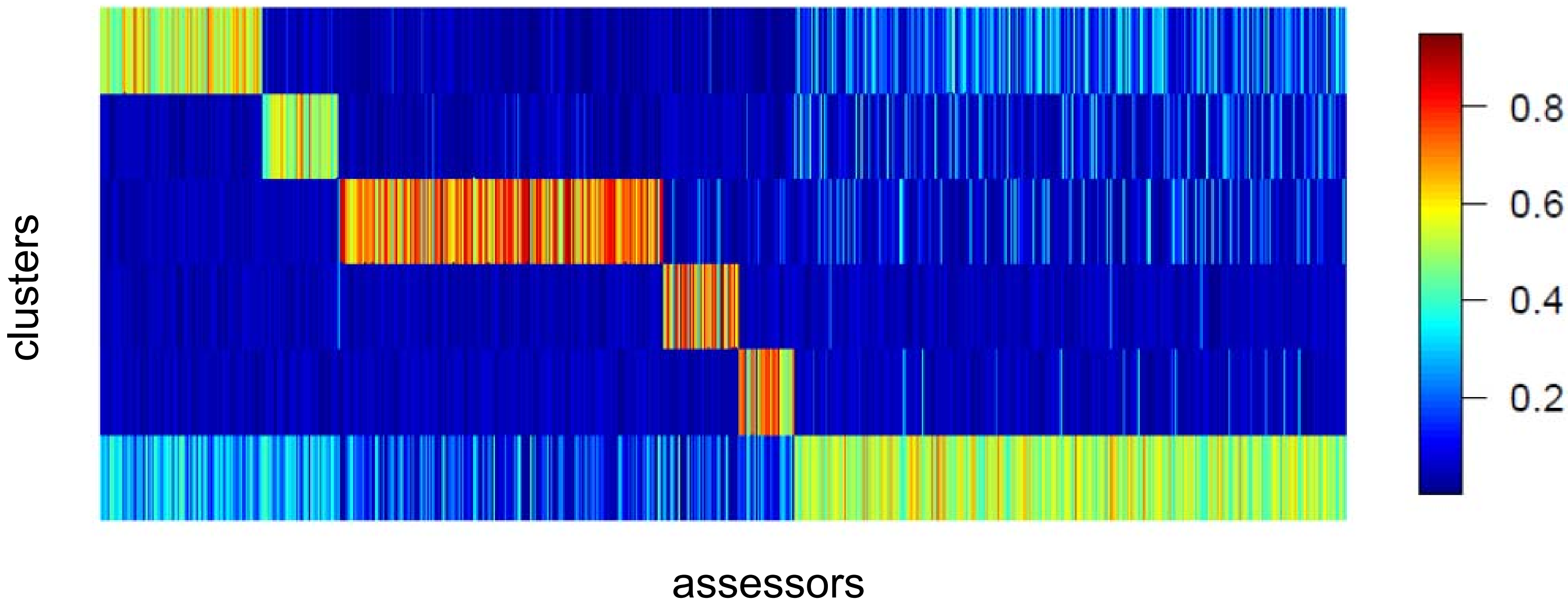
Elbow rule

MAP estimate

Cluster	$c = 1$	$c = 2$	$c = 3$
τ_c	19.08% (17.26%,20.51%)	8.97% (7.31%,10.43%)	24.03% (21.94%,25.88%)
α_c	1.56 (1.52,1.60)	1.82 (1.71,2.0)	2.9 (2.76,3.12)
	sea urchin	fatty tuna	fatty tuna
	fatty tuna	tuna	tuna
	sea eel	sea urchin	sea eel
	salmon roe	salmon roe	shrimp
	shrimp	tuna roll	tuna roll
	tuna	squid	squid
	squid	shrimp	egg
	tuna roll	sea eel	cucumber roll
	egg	egg	salmon roe
	cucumber roll	cucumber roll	sea urchin
Cluster	$c = 4$	$c = 5$	$c = 6$
τ_c	5.42% (4.55%,6.11%)	4.21% (3.95%,5.03%)	38.27% (36.14%,40.77%)
α_c	2.91 (2.78,3.15)	3.11 (2.97,3.34)	3.92 (3.77,4.15)
	shrimp	salmon roe	fatty tuna
	egg	fatty tuna	sea urchin
	squid	tuna	salmon roe
	sea eel	tuna roll	tuna
	cucumber roll	egg	shrimp
	salmon roe	shrimp	sea eel
	tuna roll	squid	tuna roll
	fatty tuna	cucumber roll	squid
	tuna	sea eel	egg
	sea urchin	sea urchin	cucumber r



- posterior probabilities for being assigned to each cluster



- most assessors have posterior probabilities concentrated on some preferred value of c, indicating a reasonably stable behaviour in the cluster assignments.
- The two clusters with the highest posterior uncertainty in assignment of assessors were c = 1 and c = 2.

Pairwise comparisons

- One observation:

$$R_1 = \{\text{coke} \prec \text{fanta}, \text{fanta} \prec \text{water}, \text{sprite} \prec \text{water}\}$$

- Transitive closure:

$$\text{tc}\,(R_1) = R_1 \cup \{\text{coke} \prec \text{water}\}$$

- Augmentation: Sample $\tilde{R}_j \in \mathcal{P}_n$ consistent with $\text{tc}(R_j)$, $j = 1, \ldots, N$

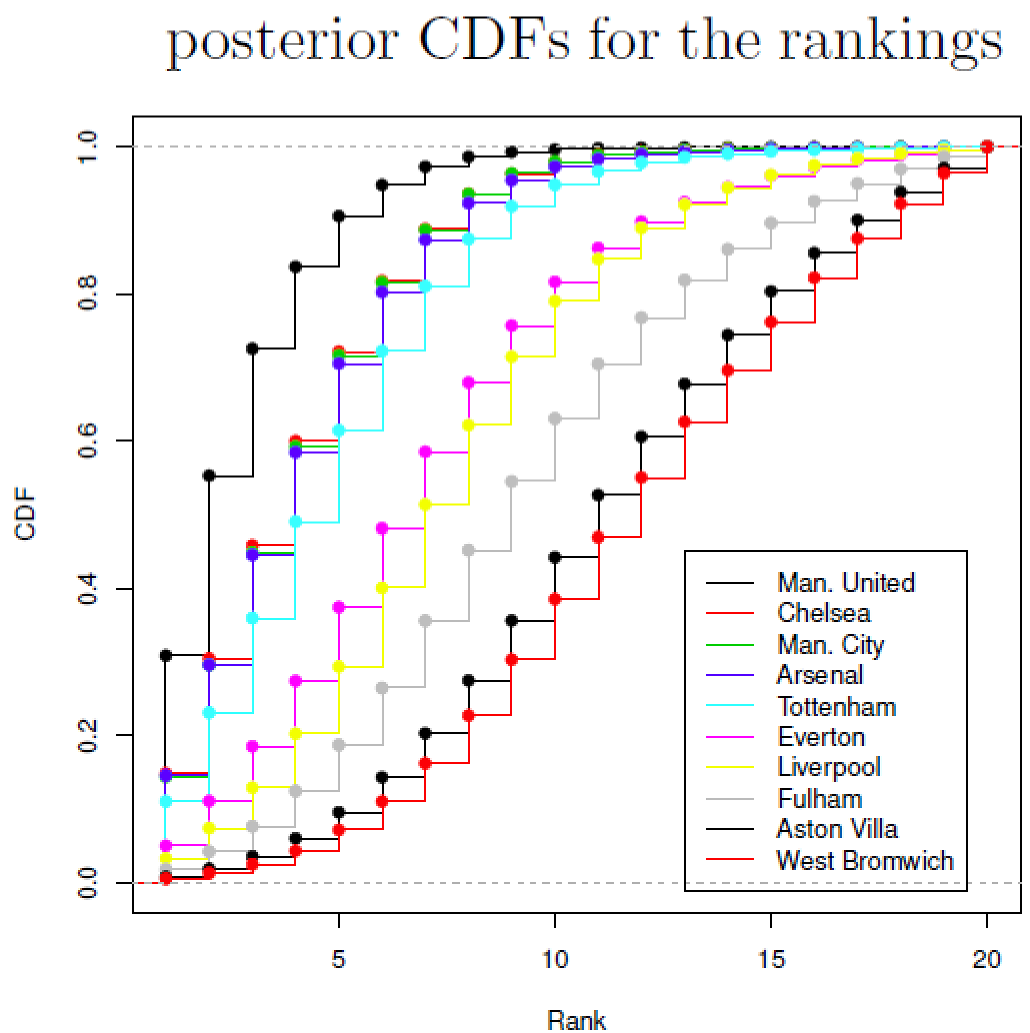
- MCMC: we need to propose augmented ranks which obey the partial ordering constraints given by the assessor.

- Assume coherent pair comparisons

- Premier League season 2010/11
- Each match a pairwise comparison between teams

cumulative probability (CP)

Official League Table				Mallows Footrule Table			
	Team	Pts.	GD		Team	CP	90 % HPDI
1	Man. United	80	+41	1	Man. United	0.31	(1,5)
2	Chelsea	71	+36	2	Chelsea	0.31	(1,8)
3	Man. City	71	+27	3	Man. City	0.45	(1,8)
4	Arsenal	68	+29	4	Arsenal	0.58	(1,8)
5	Tottenham	62	+9	5	Tottenham	0.61	(1,9)
6	Liverpool	58	+15	6	Everton	0.48	(1,13)
7	Everton	54	+6	7	Liverpool	0.51	(1,13)
8	Fulham	49	+6	8	Fulham	0.45	(2,16)
9	Aston Villa	48	-11	9	Aston Villa	0.36	(4,18)
10	Sunderland	47	-11	10	West Bromwich	0.39	(5,19)
11	West Bromwich	47	-15	11	Bolton	0.42	(6,19)
12	Newcastle	46	-1	12	Newcastle	0.50	(7,20)
13	Stoke City	46	-2	13	Sunderland	0.57	(7,20)
14	Bolton	46	-4	14	Stoke City	0.62	(8,20)
15	Blackburn	43	-13	15	Blackburn	0.61	(9,20)
16	Wigan	42	-21	16	Wigan	0.67	(8,20)
17	Wolverhampton	40	-20	17	Birmingham	0.65	(10,20)
18	Birmingham	39	-21	18	Blackpool	0.75	(10,20)
19	Blackpool	39	-23	19	Wolverhampton	0.87	(11,20)
20	West Ham	33	-27	20	West Ham	1.00	(13,20)



- perfect stochastic orderings between most of the teams
- … but not Manchester City and Chelsea

P (Team A < Team B | all data)

	Man. United	Chelsea	Man. City	Arsenal	Tottenham	Everton	Liverpool	Fulham	Aston Villa	West Bromwich	Bolton	Newcastle	Sunderland	Stoke City	Blackburn	Wigan	Birmingham	Blackpool	Wolverhampton	West Ham
Man. United	0	0.67*	0.68*	0.68*	0.73*	0.85*	0.89*	0.93*	0.96*	0.97*	0.98*	0.97*	0.98*	0.98*	0.99*	0.98*	0.99*	0.99*	1*	1*
Chelsea	0.33	0	0.5	0.51*	0.57*	0.73*	0.78*	0.85*	0.91*	0.94*	0.95*	0.94*	0.95*	0.96*	0.97*	0.96*	0.98*	0.99*	0.99*	0.99*
Man. City	0.32	0.5	0	0.5	0.57*	0.72*	0.78*	0.85*	0.91*	0.93*	0.95*	0.94*	0.94*	0.96*	0.97*	0.96*	0.98*	0.99*	0.99*	0.99*
Arsenal	0.32	0.49	0.5	0	0.56*	0.72*	0.77*	0.84*	0.91*	0.92*	0.94*	0.93*	0.94*	0.95*	0.97*	0.96*	0.98*	0.98*	0.98*	0.99*
Tottenham	0.27	0.43	0.43	0.44	0	0.66*	0.72*	0.8*	0.88*	0.9*	0.92*	0.91*	0.92*	0.94*	0.96*	0.94*	0.97*	0.97*	0.98*	0.99*
Everton	0.15	0.27	0.28	0.28	0.34	0	0.55	0.66*	0.76*	0.79*	0.82*	0.81*	0.82*	0.85*	0.88*	0.87*	0.91*	0.93*	0.93*	0.96*
Liverpool	0.11	0.22	0.22	0.23	0.28	0.45	0	0.62*	0.74*	0.77*	0.8*	0.79*	0.8*	0.83*	0.86*	0.85*	0.91*	0.92*	0.93*	0.96*
Fulham	0.07	0.15	0.15	0.16	0.2	0.34	0.38	0	0.62*	0.66*	0.7*	0.69*	0.7*	0.73*	0.78*	0.77*	0.84*	0.85*	0.86*	0.91*
Aston Villa	0.04	0.09	0.09	0.09	0.12	0.24	0.26	0.38	0	0.54*	0.57*	0.57*	0.58*	0.61*	0.67*	0.67*	0.75*	0.77*	0.78*	0.85*
West Bromwich	0.03	0.06	0.07	0.08	0.1	0.21	0.23	0.34	0.46	0	0.53*	0.54*	0.55*	0.57*	0.63*	0.63*	0.72*	0.74*	0.74*	0.82*
Bolton	0.02	0.05	0.05	0.06	0.08	0.18	0.2	0.3	0.43	0.47	0	0.51	0.51	0.54*	0.6*	0.6*	0.69*	0.71*	0.73*	0.81*
Newcastle	0.03	0.06	0.06	0.07	0.09	0.19	0.21	0.31	0.43	0.46	0.49	0	0.51*	0.53*	0.59*	0.6*	0.69*	0.71*	0.72*	0.81*
Sunderland	0.02	0.05	0.06	0.06	0.08	0.18	0.2	0.3	0.42	0.45	0.49	0.49	0	0.53*	0.59*	0.59*	0.68*	0.7*	0.71*	0.8*
Stoke City	0.02	0.04	0.04	0.05	0.06	0.15	0.17	0.27	0.39	0.43	0.46	0.47	0.47	0	0.57*	0.57	0.67*	0.68*	0.69*	0.78*
Blackburn	0.01	0.03	0.03	0.03	0.05	0.12	0.14	0.22	0.33	0.37	0.4	0.41	0.41	0.43	0	0.51	0.6*	0.62*	0.64*	0.73*
Wigan	0.02	0.04	0.04	0.04	0.06	0.13	0.15	0.23	0.33	0.37	0.4	0.4	0.41	0.43	0.49	0	0.59*	0.61*	0.62*	0.72*
Birmingham	0.01	0.02	0.02	0.02	0.03	0.09	0.09	0.16	0.25	0.28	0.31	0.31	0.32	0.33	0.4	0.41	0	0.52*	0.52*	0.64*
Blackpool	0.01	0.01	0.01	0.02	0.03	0.07	0.08	0.15	0.23	0.26	0.29	0.29	0.3	0.32	0.38	0.39	0.48	0	0.51	0.62*
Wolverhampton	0	0.01	0.01	0.02	0.02	0.07	0.07	0.14	0.22	0.26	0.27	0.28	0.29	0.31	0.36	0.38	0.48	0.49	0	0.61*
West Ham	0	0.01	0.01	0.01	0.01	0.04	0.04	0.09	0.15	0.18	0.19	0.19	0.2	0.22	0.27	0.28	0.36	0.38	0.39	0

Table S7: Stochastic orderings in the Premier League example. Each matrix element (i, j) denotes the posterior probability that team A_i has a lower rank than team A_j. If the number is marked with an asterisk, it means that A_i stochastically dominates A_j.

Preference Prediction

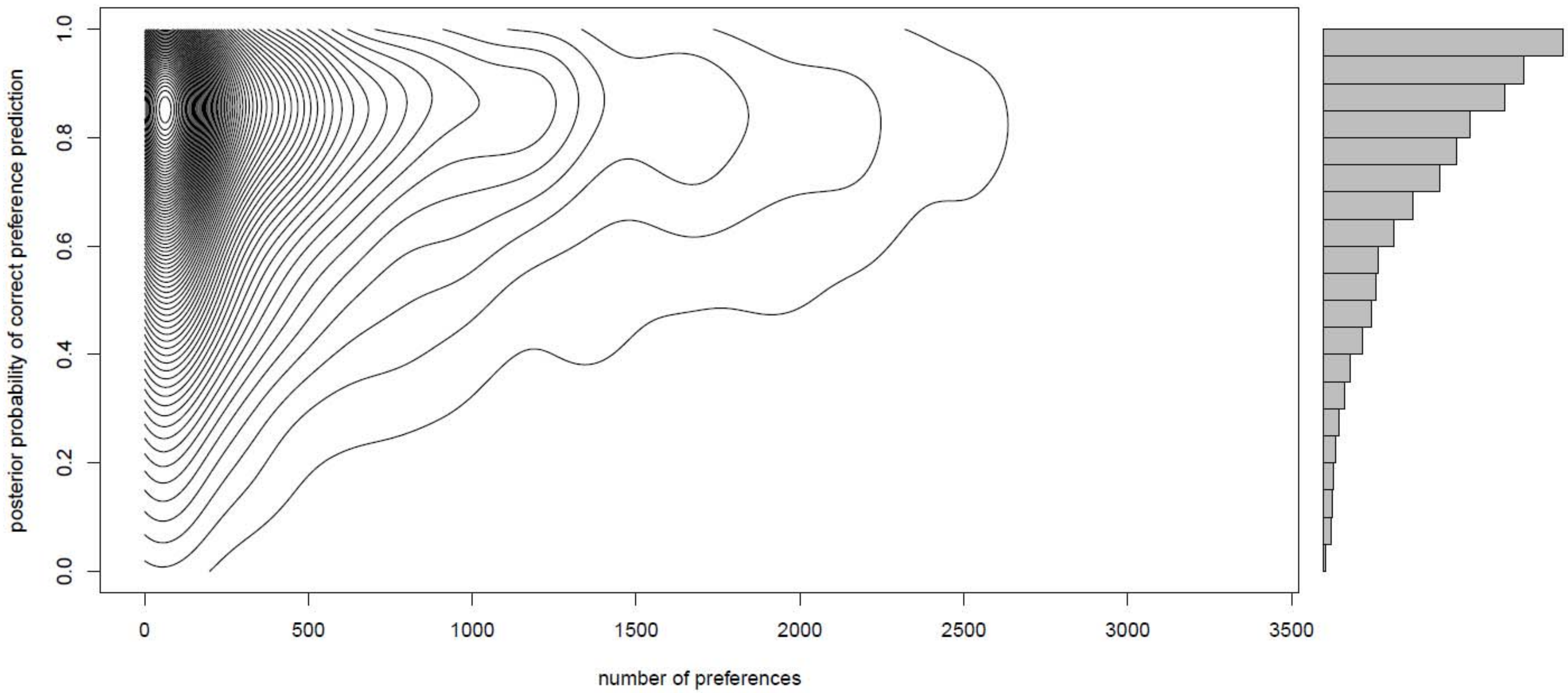
movielens
helping you find the *right* movies

- N=5891 assessors (users), n=200 movies
  Mean number of movies rated per user = 30.2
- Ratings transformed to pair comparisons (as in Lu & Boutilier 2011)
- 14 classes of users (age and gender) – for simplicity fixed, in real application would be estimated
- Normalising constant approximated as in Mukherjee 2013, as importance sampling inefficient with n=200
- $P(\tilde{\mathbf{R}}_j | \text{all data})$ is the posterior predicted probability of the full ranking for assessor j, consistent with given preferences and relative to the class j belongs to.
- **Personalised recommendation**.

- To test method, we discarded one rated movie per user
- Use all other data to estimate $P(\tilde{\mathbf{R}}_j|\text{all data})$
- Read off posterior probability of the given (but hidden) preference
- Median such probability over all assessors = 0.812
- If we decide to predict the preference between two movies by taking the preference with posterior predictive probability >0.5, then we make an error in 12.7% of cases.

- Contour plot of the posterior probability of correctly predicting the discarded preference vs. the number of preferences stated by the assessor.

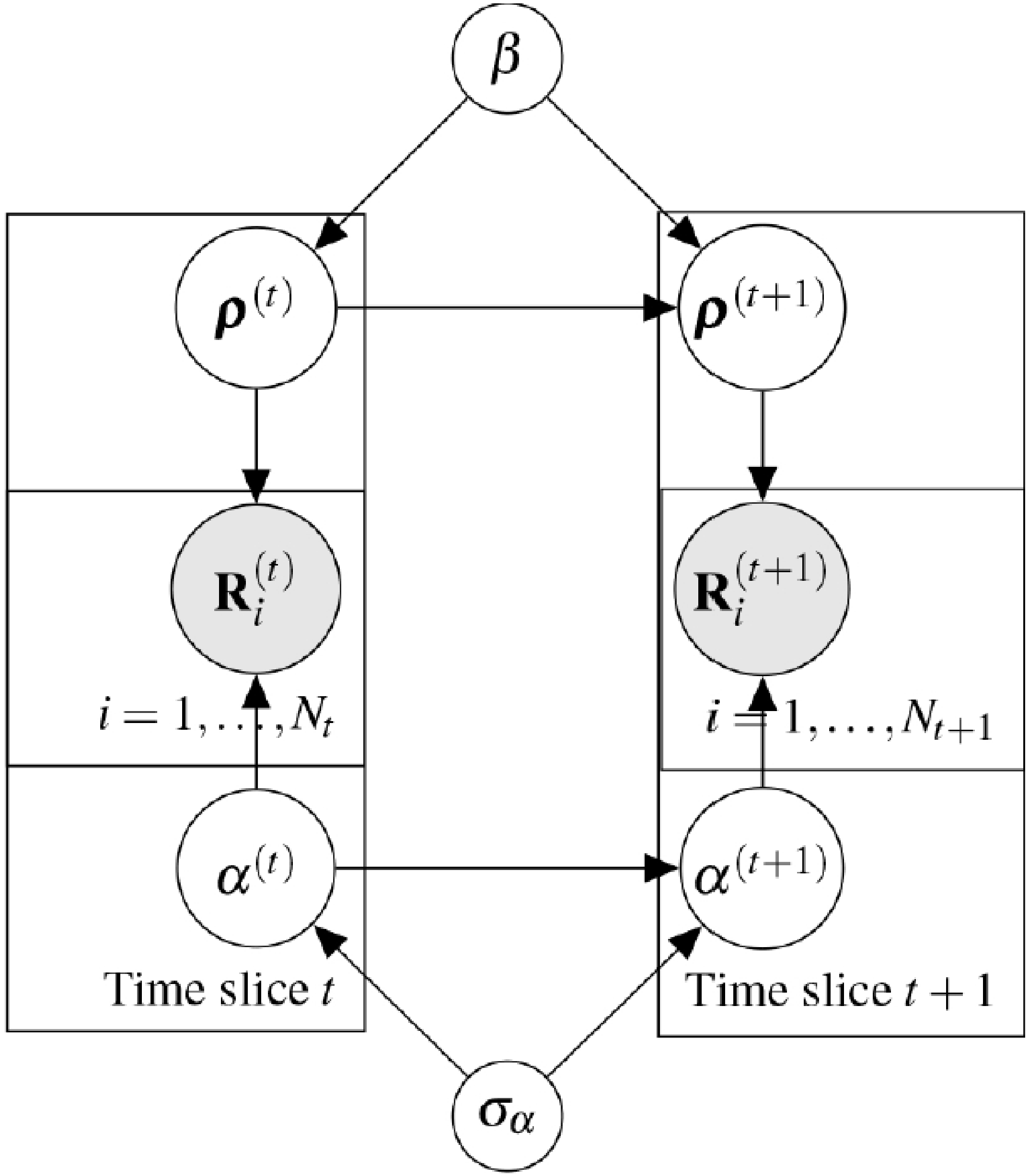
- Histogram shows the marginal posterior probability of correct preference prediction.



Time-dependent ranks

Preferences often develop over time, e.g.,

- New products enter the market, while others go out of fashion
- Political views
- Athletes and sports teams

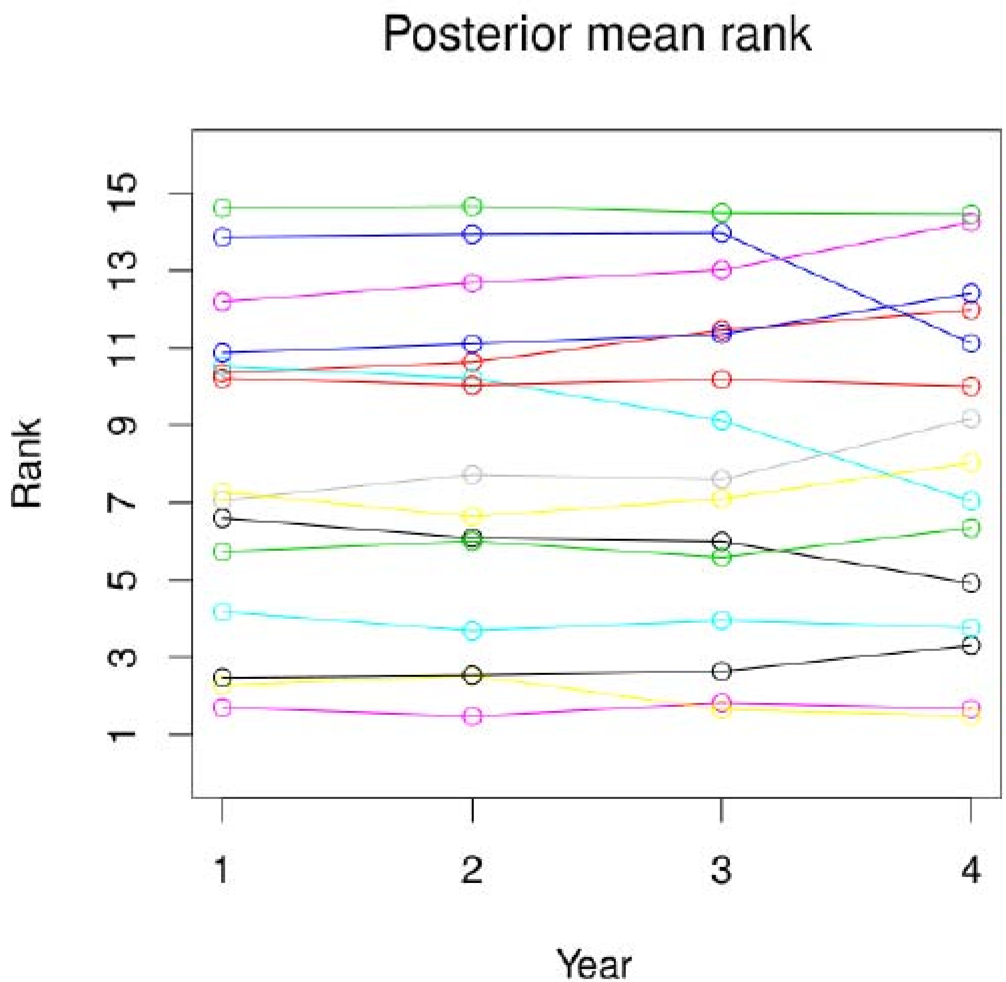


β
$\boldsymbol{\rho}^{(t)}$
$\boldsymbol{\rho}^{(t+1)}$
$\mathbf{R}_i^{(t)}$
$\mathbf{R}_i^{(t+1)}$
$i = 1, \ldots, N_t$
$i = 1, \ldots, N_{t+1}$
$\boldsymbol{\alpha}^{(t)}$
$\boldsymbol{\alpha}^{(t+1)}$
Time slice t
Time slice $t+1$
σ_α

We model the transition between time steps with the Mallows model

$$P\left(\boldsymbol{\rho}^{(t)}|\boldsymbol{\rho}^{(t-1)},\beta\right) = Z_n\left(\beta\right)^{-1}\exp\left\{\frac{-\beta}{n}d\left(\boldsymbol{\rho}^{(t)},\boldsymbol{\rho}^{(t-1)}\right)\right\}1_{\mathcal{P}_n}\left(\boldsymbol{\rho}^{(t)}\right)$$

- We followed 15 high school students over four years
- Between 4 and 8 math tests per year



Conclusions

- Our Bayesian Mallows model handles
  - Partial data
  - Pairwise comparisons
  - Clustering
  - Time-dependent data
  - Any distance measure!

Future developments

- Covariances of assessors
- Covariances of items
- Not coherent pair comparisons
- Comparison of distances
- ABC instead of importance sampling
- Items changing over time
- MCMC scales poorly with number of items

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

Diaconis, P. and R. L. Graham (1977). Spearman's footrule as a measure of disarray. *Journal of the Royal Statistical Society. Series B (Methodological) 39*(2), 262–268.

Lu, T. and C. Boutilier (2011). Learning Mallows models with pairwise preferences. In L. Getoor and T. Scheffer (Eds.), *Proceedings of the 28th International Conference on Machine Learning*, ICML '11, New York, NY, USA, pp. 145–152. ACM.

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

Marden, J. I. (1995). *Analyzing and Modeling Rank Data*, Volume 64 of *Monographs on Statistics and Applied Probability*. Cambridge, MA, USA: Chapman & Hall.

Mukherjee, S. (2013). Estimation of parameters in non uniform models on permutations. arXiv:1307.0978 [math.PR].

- Visual inspection: 'Footrule'$=\ell_1$-distance