A multiresolution framework for the statistical analysis of ranking data

Eric Sibony   Stéphan Clémençon   Jérémie Jakubowicz
Anna Korba

LTCI UMR 5141, Telecom ParisTech/CNRS

CIRM Workshop on Statistical Learning
February 2 2016

Outline

Many modern systems gather feedback data about the preferences of their users

Many modern systems gather feedback data about the preferences of their users

Many modern systems gather feedback data about the preferences of their users

Many modern systems gather feedback data about the preferences of their users

This ranking data only asks to be analyzed.

$\Rightarrow$ Over the past 15 years, the statistical analysis of ranking data has become a subfield of the machine learning literature.

However, ranking data also arise in many other applications

Example 1: Elections

Set of candidates $\{A, B, C, D\}$. A voter can give for instance

- her favorite candidate, e.g. $B \succ A, C, D$.
- a full ordering of the candidates, e.g. $B \succ D \succ A \succ C$

The collection of ballots in an election is a **dataset of rankings**.
$\Rightarrow$ How to elect the winner(s)?

Borda-Condorcet debate since the 18^{th} century

Jean-Charles de Borda

Nicolas de Condorcet

However, ranking data also arise in many other applications

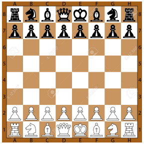
Example 2: Competitions

Set of participants $\{1, \ldots, n\}$.
The results of a game can be for instance

- Victory of i against j: $i \succ j$
  (e.g. in football, chess, ...)
- Ranking of the participants of the game:
  $i_1 \succ \cdots \succ i_k$
  (e.g. races, video games, ...)

The results of the games of a competition form a **dataset of rankings**.

$\Rightarrow$ What is the ranking of the competition?



However, ranking data also arise in many other applications

Example 3: Surveys

Set of items $\{1, \ldots, n\}$.
A respondent can be asked to give for instance

- Her top-3 items: $i_1, i_2, i_3 \succ$ *the rest*
- Her top-3 ranking of items: $i_1 \succ i_2 \succ i_3 \succ$ *the rest*

The answers to a survey constitute a **dataset of rankings**.

$\Rightarrow$ How to summarize the results?
$\Rightarrow$ How to segment respondents based on their answers?

The analysis of ranking data spreads over many fields of the scientific literature

- Machine learning
- Social choice theory
- Economics
- Psychology
- Operational Research
- Artificial intelligence

Many efforts to bring them together

NIPS 2001	New Methods for Preference Elicitation
NIPS 2002	Beyond Classification and Regression
KI 2003	Preference Learning
NIPS 2004	Learning with Structured Outputs
NIPS 2005	Learning to Rank
IJCAI 2005	Advances in Preference Handling
SIGIR 07-10	Learning to Rank for Information Retrieval
ECML/PKDD 08-10	Preference Learning
NIPS 09	Advances in Ranking
AIM Workshop 2010	The Mathematics of Ranking
NIPS 2011	Choice Models and Preference Learning
EURO 09-16	Special track on Preference Learning
ECAI 2012	Preference Learning
DA2PL 2012,2014	From Decision Analysis to Preference Learning
Dagstuhl 2014	Seminar on Preference Learning
NIPS 2014	Analysis of Rank Data

Outline

Definitions

Set of items $[\![n]\!] := \{1, \ldots, n\}$

Definition (Ranking)

A ranking is a strict partial order $\prec$ over $[\![n]\!]$, *i.e.* a binary relation satisfying the following properties:

- **Irreflexivity** For all $a \in [\![n]\!]$, $a \not\prec a$
- **Transitivity** **For all $\mathbf{a}, \mathbf{b}, \mathbf{c} \in [\![\mathbf{n}]\!]$, if $\mathbf{a} \prec \mathbf{b}$ and $\mathbf{b} \prec \mathbf{c}$ then $\mathbf{a} \prec \mathbf{c}$**
- **Asymmetry** For all $a, b \in [\![n]\!]$, if $a \prec b$ then $b \not\prec a$

Shortcut notations:

- $a \succ b \succ c$ instead of $(a \succ b, a \succ c, b \succ c)$
- $a \succ b, c$ instead of $(a \succ b, a \succ c)$

Main types of rankings

- **Full ranking.** All the items are ranked, without ties

$$a_1 \succ \cdots \succ a_n$$

- **Partial ranking.** All the items are ranked, with ties

$$a_{1,1}, \ldots, a_{1,n_1} \succ \cdots \succ a_{r,1}, \ldots, a_{r,n_r} \quad \text{with} \quad \sum_{i=1}^{r} n_i = n$$

- **Incomplete ranking.** Only a subset of items are ranked, without ties

$$a_1 \succ \cdots \succ a_k \quad \text{with} \quad k < n$$

One can further consider incomplete and partial rankings.

General setting

Perform some task on a dataset of N rankings $\mathcal{D}_N = (\prec_1, \dots, \prec_N)$.

Examples

- **Top-1 recovery:** Find the "most preferred" item in $\mathcal{D}_N$
  e.g. Output of an election
- **Aggregation:** Find a full ranking that "best summarizes" $\mathcal{D}_N$
  e.g. Ranking of a competition
- **Clustering:** Split $\mathcal{D}_N$ into clusters
  e.g. Segment customers based on their answers to a survey
- **Prediction:** Predict the outcome of a missing pairwise comparison in a ranking $\prec$
  e.g. In a recommendation setting

Detailed example: analysis of full rankings

Notation.

- The full ranking $a_1 \succ \cdots \succ a_n$ is denoted by $a_1 \dots a_n$
- Also seen as the permutation σ that maps an item to its rank:

$$a_1 \succ \cdots \succ a_n \quad \Leftrightarrow \quad \sigma \in \mathfrak{S}_n \text{ such that } \sigma(a_i) = i$$

$\mathfrak{S}_n$: set of permutations of $[\![n]\!]$, the symmetric group.

Probabilistic Modeling. The dataset is a collection of random permutations drawn IID from a probability distribution p over $\mathfrak{S}_n$:

$$\mathcal{D}_N = (\Sigma^{(1)}, \dots, \Sigma^{(N)}) \qquad \text{with} \qquad \Sigma^{(i)} \sim p$$

p is called a ranking model.

Detailed example: analysis of full rankings

Example, dataset from [Croon, 1989]

After the fall of the Berlin wall a survey of German citizens was conducted where they were asked to rank four political goals

1. Maintain order
2. Give people more say in government
3. Fight rising prices
4. Protect freedom of speech

Detailed example: analysis of full rankings

Example, dataset from [Croon, 1989]

They collected 2,262 answers

Ranking	Answers	Ranking	Answers
1234	137	3124	330
1243	29	3142	294
1324	309	3214	117
1342	255	3241	69
1423	52	3412	70
1432	93	3421	34
2134	48	4123	21
2143	23	4132	30
2314	61	4213	29
2341	55	4231	52
2413	33	4312	35
2431	39	4321	27

Detailed example: analysis of full rankings

Questions

- How to analyze a dataset of permutations $\mathcal{D}_N = (\Sigma^{(1)}, \ldots, \Sigma^{(N)})$?
- How to characterize the variability?
- What can be inferred?

Detailed example: analysis of full rankings

Notation:

$$\sigma = \begin{pmatrix} 1 & 2 & \dots & n \\ \sigma(1) & \sigma(2) & \dots & \sigma(n) \end{pmatrix} = (\sigma(1), \sigma(2), \dots, \sigma(n))$$

A random permutation Σ can be seen as a random vector

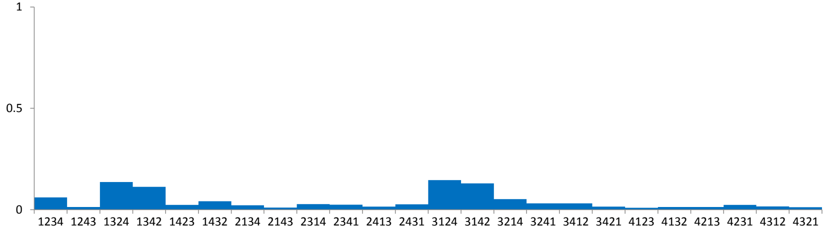
$$(\Sigma(1), \dots, \Sigma(n)) \in \mathbb{R}^n$$

But

- The random variables $\Sigma(1), \dots, \Sigma(n)$ are highly dependent
- The sum $\Sigma + \Sigma'$ is not a random permutation
  $\Rightarrow$ No law of large numbers nor central limit theorem on $\mathfrak{S}_n$
- No natural notion of variance for Σ

Detailed example: analysis of full rankings

The set of permutations $\mathfrak{S}_n$ is finite, compute the histogram:

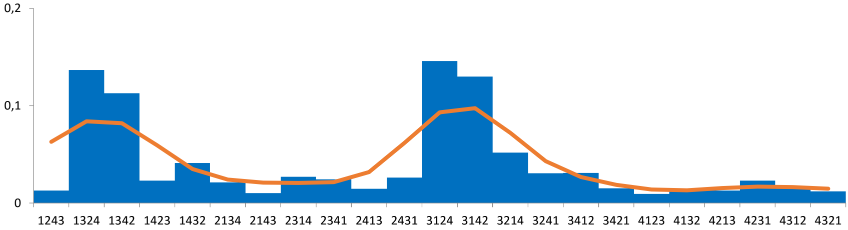


But

- Exploding cardinality: $|\mathfrak{S}_n| = n!$. *E.g.* $20! = 2.4 \times 10^{18}$
  $\Rightarrow$ Few statistical relevance

Detailed example: analysis of full rankings

Apply a method from p.d.f. estimation (e.g. kernel density estimation):



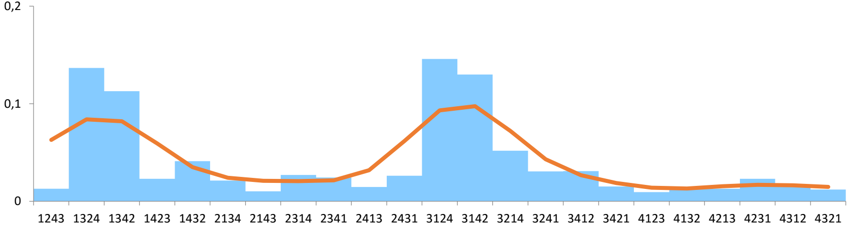
But

- No canonical ordering of the rankings

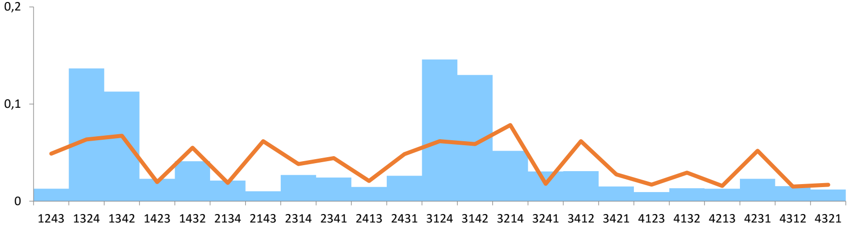
Detailed example: analysis of full rankings

No canonical ordering of the rankings

Lexicographical ordering



Random ordering



Detailed example: analysis of full rankings

More generally: many possible distances on $\mathfrak{S}_n$

- Kendall's tau distance

$$d(\sigma,\pi) = \sum_{1\leq i<j\leq n} \mathbb{I}\{\sigma \text{ and } \pi \text{ disagree on } \{i,j\}\}$$

- Spearman's rho distance (l^2 norm)

$$d(\sigma,\pi) = \left(\sum_{i=1}^{n} (\sigma(i)-\pi(i))^2\right)^{1/2}$$

- Footrule distance (l^1 norm)

$$d(\sigma,\pi) = \sum_{i=1}^{n} |\sigma(i)-\pi(i)|$$

- Many others: Hamming, Cayley, Ulam, ...

Detailed example: analysis of full rankings

More generally: many possible graph structures on $\mathfrak{S}_n$

- Adjacent transpositions

$$\sigma \text{ and } \pi \text{ are neighbors if } \pi = (i\ i+1)\sigma \text{ with } 1 \leq i \leq n-1$$

- All transpositions

$$\sigma \text{ and } \pi \text{ are neighbors if } \pi = (i\ j)\sigma \text{ with } 1 \leq i \neq j \leq n$$

- Star graph

$$\sigma \text{ and } \pi \text{ are neighbors if } \pi = (1\ k)\sigma \text{ with } 2 \leq k \leq n$$

- ...

Detailed example: analysis of full rankings

More generally: many possible embeddings of $\mathfrak{S}_n$

- Permutation matrices

$$\mathfrak{S}_n \to \mathbb{R}^{n \times n}, \quad \sigma \mapsto P_\sigma \quad \text{with } P_\sigma(i,j) = \mathbb{I}\{\sigma(i) = j\}$$

- embedding in a sphere
- embedding as angles
- ...

Detailed example: analysis of full rankings

Exploit any of the combinatorial or algebraic properties of $\mathfrak{S}_n$

But

- Ranking data are very natural for human beings
  $\Rightarrow$ Statistical modeling should capture some interpretable structure

Detailed example: analysis of full rankings

“Parametric” approach

- Fit a predefined generative model on the data
- Analyze the data through that model
- Infer knowledge with respect to that model

“Nonparametric” approach

- Choose a structure on $\mathfrak{S}_n$
- Analyze the data with respect to that structure
- Infer knowledge through a “regularity” assumption

Detailed example: analysis of full rankings

Parametric approach - classic models

- Mallows model [Mallows, 1957]

$$p(\sigma) = Ce^{-\gamma d(\sigma_0, \sigma)} \qquad \text{with } \sigma \in \mathfrak{S}_n \text{ and } \gamma \in \mathbb{R}^+$$

- Plackett-Luce model [Luce, 1959], [Plackett, 1975]

$$p(\sigma) = \prod_{i=1}^{n} \frac{w_{\sigma_i}}{\sum_{j=i}^{n} w_{\sigma_j}} \qquad \text{with } w_i \in \mathbb{R}^+$$

- Thurstone model [Thurstone, 1927]

$$p(\sigma) = \int_{x_1 > \cdots > x_n} \prod_{i=1}^{n} f_i(x_i) dx_i \qquad \text{with } f_i \text{ p.d.f. on } \mathbb{R}$$

Detailed example: analysis of full rankings

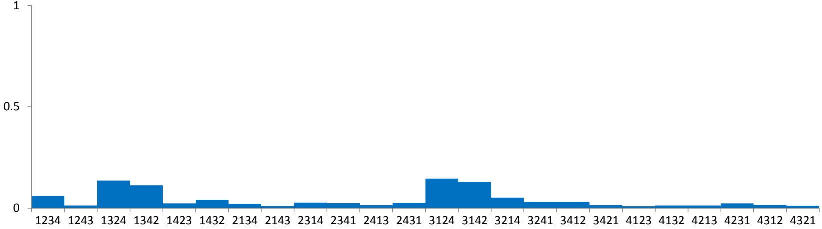
Examples of nonparametric approaches

- Distance-based modeling
- Independence modeling
- Embedding in euclidean space
- Pairwise decomposition
- Sparsity assumption
- Sampling-based models
- Algebraic toric models
- **Harmonic analysis**

Outline

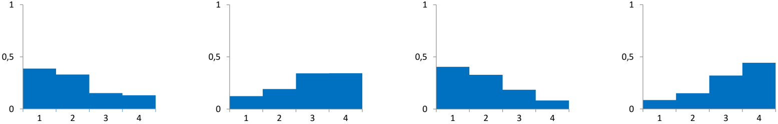
Ranking models can be analyzed through their marginals

Law of Σ



$\Downarrow$

Law of $\Sigma(1)$ Law of $\Sigma(2)$ Law of $\Sigma(3)$ Law of $\Sigma(4)$



Ranking models can be analyzed through their marginals

The law of $\Sigma(i)$ for $i \in [\![n]\!]$ is naturally given by

$$\mathbb{P}[\Sigma(i) = j] = \sum_{\sigma \in \mathfrak{S}_n,\ \sigma(i)=j} p(\sigma)$$

It is called a marginal of p of **order 1**.

Ranking models can be analyzed through their marginals

The law of $\Sigma(i)$ for $i \in [\![n]\!]$ is naturally given by

$$\mathbb{P}[\Sigma(i) = j] = \sum_{\sigma \in \mathfrak{S}_n,\ \sigma(i)=j} p(\sigma)$$

It is called a marginal of p of **order 1**.

Marginals of order 2

- Unordered: law of $\Sigma(\{i_1, i_2\})$

$$\mathbb{P}[\Sigma(\{i_1, i_2\}) = \{j_1, j_2\}] = \sum_{\sigma \in \mathfrak{S}_n,\ \sigma(\{i_1,i_2\})=\{j_1,j_2\}} p(\sigma)$$

- Ordered: law of $\Sigma((i_1, i_2))$

$$\mathbb{P}[\Sigma((i_1, i_2)) = (j_1, j_2)] = \sum_{\sigma \in \mathfrak{S}_n,\ \sigma((i_1,i_2))=(j_1,j_2)} p(\sigma)$$

Ranking models can be analyzed through their marginals

In general, one can consider the law of $(\Sigma(A_1), \ldots, \Sigma(A_r))$, where $(A_1, \ldots, A_r)$ is an ordered partition of $[\![n]\!]$:

$$\mathbb{P}\left[\Sigma(A_1) = B_1, \ldots, \Sigma(A_r) = B_r\right] = \sum_{\substack{\sigma \in \mathfrak{S}_n \\ \sigma(A_1)=B_1,\ldots,\sigma(A_r)=B_r}} p(\sigma)$$

It is called a marginal of order $\lambda := (|A_1|, \ldots, |A_r|)$

Remark: λ is a partition of n, i.e. $\lambda \in \mathbb{N}^r$ is such that $\lambda_1 \geq \cdots \geq \lambda_r \geq 1$ and $\sum_{i=1}^r \lambda_i = n$

	$\lambda = (n-1, 1)$	Order 1
Example	$\lambda = (n-2, 2)$	Order 2, unordered
	$\lambda = (n-2, 1, 1)$	Order 2, ordered

Ranking models can be analyzed through their marginals

Let $M^{\lambda} p$ denote all the marginals of p of order λ.

Analyzing the marginals has two main purposes

Ranking models can be analyzed through their marginals

Let $M^\lambda p$ denote all the marginals of p of order λ.

Analyzing the marginals has two main purposes

1. For each λ, $M^\lambda p$ **focus on some part of the variability of p**

Ranking models can be analyzed through their marginals

Let $M^{\lambda}p$ denote all the marginals of p of order λ.

Analyzing the marginals has two main purposes

1. For each λ, $M^{\lambda}p$ **focus on some part of the variability of p**
2. In a statistical setting, instead of considering the $n!$-dimensional empirical estimator

$$\widehat{p}_N = \frac{1}{N}\sum_{t=1}^{N}\delta_{\Sigma(t)},$$

analyzing $M^{\lambda}\widehat{p}_N$ allows to **reduce the dimension**

Example: $M^{(n-1,1)}\widehat{p}_N \in \mathbb{R}^{n\times n} \quad \Rightarrow \quad \dim = n^2$

Ranking models can be analyzed through their marginals

What λ to choose?

Marginals have a structure

Marginals contain nested part of information

Example: Order $(n-2,1,1)$ marginals induce order $(n-1,1)$ marginals

For $i,j \in [\![n]\!]$, one has

$$\mathbb{P}[\Sigma(i)=j] = \sum_{\substack{j'=1 \\ j'\neq j}}^{n} \mathbb{P}[\Sigma(i)=j, \Sigma(i')=j'] \quad \text{for any } i' \neq i$$

$\Rightarrow$ There is a linear operator $\Phi : \mathbb{R}^{n(n-1)\times n(n-1)} \to \mathbb{R}^{n\times n}$ such that

$$\Phi : \quad M^{(n-2,1,1)}p \quad \mapsto \quad M^{(n-1,1)}p$$

$\Rightarrow$ The knowledge of $M^{(n-2,1,1)}p$ induces the knowledge of $M^{(n-1,1)}p$

Marginals contain nested part of information

More generally, one can show the following result

Proposition

λ-marginals are nested according to an order $\trianglelefteq$ on partitions of n. In particular:

$$(n) \trianglelefteq (n-1,1) \trianglelefteq (n-2,2) \trianglelefteq (n-2,1,1) \trianglelefteq \cdots \trianglelefteq (1,\dots,1)$$

where

- (n) *is order 0:* $M^{(n)}p = \sum_{\sigma \in \mathfrak{S}_n} p(\sigma)$
- $(1,\dots,1)$ *is highest order:* $M^{(1,\dots,1)}p = p$

Marginals contain nested part of information

		Dimension
Ranking model on $\mathfrak{S}_n$	p	$n!$
$\cup$		
$\cdots$	$\downarrow$	
$\cup$		
Ordered order 2 marginals	$M^{(n-2,1,1)}p$	$n^2(n-1)^2$
$\cup$	$\downarrow$	
Unordered order 2 marginals	$M^{(n-2,2)}p$	$\binom{n}{2}^2$
$\cup$	$\downarrow$	
Order 1 marginals	$M^{(n-1,1)}p$	n^2
$\cup$	$\downarrow$	
Order 0 marginal	$M^{(n)}p$	1

Ranking models can be analyzed through their marginals

What λ to choose?

With Fourier analysis

Fourier analysis on the symmetric group in a nutshell

Introduced by Persi Diaconis in [Diaconis, 1988]
Many developments since then (e.g. [Huang et al., 2009], [Kondor and Barbosa, 2010], [Kakarala, 2011]).

Definition (Fourier transform)

The Fourier transform of a function $f : \mathfrak{S}_n \to \mathbb{R}$ is defined by

$$\mathcal{F} : f \mapsto \left(\widehat{f}(\lambda)\right)_\lambda \qquad \text{with} \qquad \widehat{f}(\lambda) = \sum_{\sigma \in \mathfrak{S}_n} f(\sigma)\rho_\lambda(\sigma)$$

where $\sigma \mapsto \rho_\lambda(\sigma)$ is an *irreducible representation* of $\mathfrak{S}_n$.

Analogy with Fourier series

$\sigma \mapsto \rho_\lambda(\sigma)$	is the equivalent of	$e_k : x \mapsto e^{2i\pi kx}$
$\widehat{f}(\lambda)$	is the equivalent of	$\widehat{f}(k) = \langle f, e_k \rangle$

Fourier transform on the symmetric group in a nutshell

Some differences with the classic Fourier transform

- Fourier coefficients are matrices: $\widehat{f}(\lambda) \in \mathbb{R}^{d_\lambda \times d_\lambda}$
- "Frequencies" λ are partitions of n (no natural interpretation)

Satisfies though some classic properties

- Parseval identity
- Inverse Fourier transform
- Turns convolution into (matrix) product
- Fast Fourier Transform

$\widehat{f}(\lambda)$ localizes specific information the λ-marginals of f

We recall that $\trianglelefteq$ is the order on partitions of n such that

$$\lambda \trianglerighteq \mu \quad \Leftrightarrow \quad M^\lambda f \text{ induces } M^\mu f$$

Theorem (Young's rule, informal)

For any λ,

$$M^\lambda p \quad \text{``$=$''} \quad \widehat{p}(\lambda) + F((M^\mu p)_{\mu \triangleleft \lambda})$$

where F is some function

$\widehat{f}(\lambda)$ localizes specific information the λ-marginals of f

Example

- $\widehat{p}(n)$ localizes information specific to order 0 marginal:

$$\widehat{p}(n) = M^{(n)} p \qquad \left(:= \sum_{\sigma \in \mathfrak{S}_n} p(\sigma) \right)$$

- $\widehat{p}(n-1,1)$ localizes information specific to order 1 marginals:

$$M^{(n-1,1)} p \quad \text{``}=\text{''} \quad \widehat{p}(n-1,1) + F(M^{(n)} p)$$

- $\widehat{p}(n-2,2)$ localizes information specific to unordered order 2 marginals:

$$M^{(n-2,2)} p \quad \text{``}=\text{''} \quad \widehat{p}(n-2,2) + F(M^{(n-1,1)} p + M^{(n)} p)$$

Analysis through the Fourier transform

	Marginal	Fourier coefficient
Ranking model on $\mathfrak{S}_n$	p	$\widehat{p}(1,\dots,1)$
$\cup$		
$\dots$	$\downarrow$	
$\cup$		
Ordered order 2 marginals	$M^{(n-2,1,1)}p$	$\widehat{p}(n-2,1,1)$
$\cup$	$\downarrow$	
Unordered order 2 marginals	$M^{(n-2,2)}p$	$\widehat{p}(n-2,2)$
$\cup$	$\downarrow$	
Order 1 marginals	$M^{(n-1,1)}p$	$\widehat{p}(n-1,1)$
$\cup$	$\downarrow$	
Order 0 marginal	$M^{(n)}p$	$\widehat{p}(n)$

Analysis through the Fourier transform

The Fourier transform allows to measure how much information is contained in each order λ:

$$\|f\|^2 = \sum_{\lambda} d_\lambda \|\widehat{f}(\lambda)\|^2 \qquad \text{(Parseval identity)}$$

Example on the dataset from [Croon, 1989]

λ	(4)	$(3,1)$	$(2,2)$	$(2,1,1)$	$(1,1,1,1)$
$\frac{d_\lambda \|\widehat{f}(\lambda)\|^2}{\|f\|^2}$	49%	**33%**	**17%**	1%	0%

The analysis can then go much further

Outline

λ-marginals localize absolute rank information.

Rank information

Permutation $\begin{pmatrix} 1 & 2 & 3 & 4 & 5 \\ 2 & 5 & 4 & 3 & 1 \end{pmatrix} \leftrightarrow$ Ranking $5 \succ 1 \succ 4 \succ 3 \succ 2$

Absolute rank information

- What is the rank $\sigma(3)$ of item 3? 4
- What item $\sigma^{-1}(2)$ is ranked at 2^{nd} position? 1
- What are the ranks $\sigma(\{2,4,5\})$ of items $\{2,4,5\}$? $\{5,3,1\}$

Relative rank information

- How are items 1 and 3 relatively ordered? $1 \succ 3$
- How are the items of the subset $\{2,4,5\}$ relatively ordered? $5 \succ 4 \succ 2$

Rank information

$$\text{Permutation } \sigma \quad \leftrightarrow \quad \text{Ranking } \sigma^{-1}(1) \succ \cdots \succ \sigma^{-1}(n)$$

Absolute rank information

- What is the rank $\sigma(i)$ of item i?
- What item $\sigma^{-1}(j)$ is ranked at j^{th} position?
- What are the ranks $\sigma(\{i,j,k\})$ of items $\{i,j,k\}$?

Relative rank information

- How are items a and b relatively ordered?
- How are the items of the subset A relatively ordered?

Rank information

$$\text{rnd Permutation } \Sigma \quad \leftrightarrow \quad \text{rnd Ranking } \Sigma^{-1}(1) \succ \cdots \succ \Sigma^{-1}(n)$$

Absolute rank information

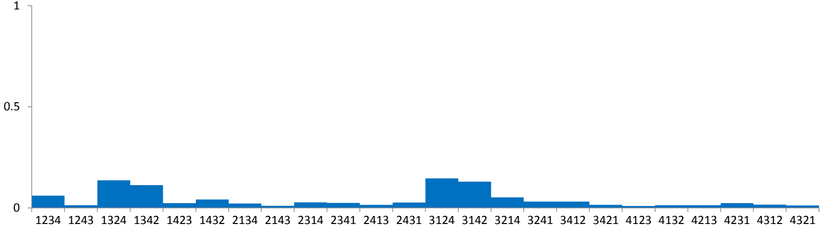
- What is the law of the rank $\Sigma(i)$ of item i ?
- What is the law of the item $\Sigma^{-1}(j)$ ranked at j^{th} position?
- What is the law of the ranks $\Sigma(\{i, j, k\})$ of items $\{i, j, k\}$?

Relative rank information

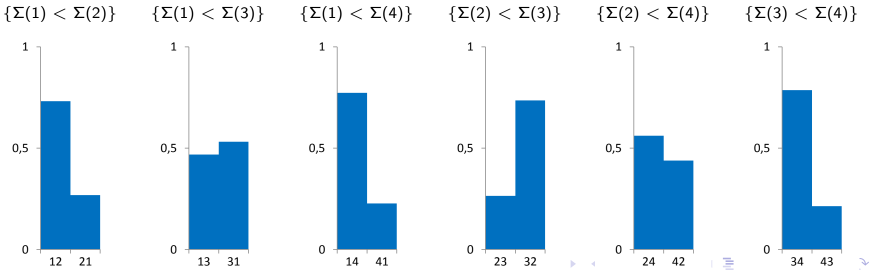
- What is the probability $\mathbb{P}[\Sigma(a) < \Sigma(b)]$ that a is ranked higher than b?
- What is the law of the ranking $\Sigma_{|A}$ induced by Σ on the subset A?

Relative marginals provide a different point of view

Law of Σ



$$\Downarrow$$



Relative marginals provide a different point of view

Definition (Induced ranking)

For $\sigma \in \mathfrak{S}_n$ and $A \subset [\![n]\!]$ with $|A| \geq 2$, we denote by $\sigma_{|A}$ the ranking induced by σ on the items of A.

$$\text{e.g.} \quad \sigma = 24153, \quad \sigma_{|\{2,3,5\}} = 253$$

Definition (Relative marginals)

The marginal of p on a subset $A \subset [\![n]\!]$ with $|A| \geq 2$ is the law of the ranking $\Sigma_{|A}$, given by

$$M_A p = \mathbb{P}[\Sigma_{|A} = \pi] = \sum_{\sigma \in \mathfrak{S}_n,\ \sigma_{|A} = \pi} p(\sigma)$$

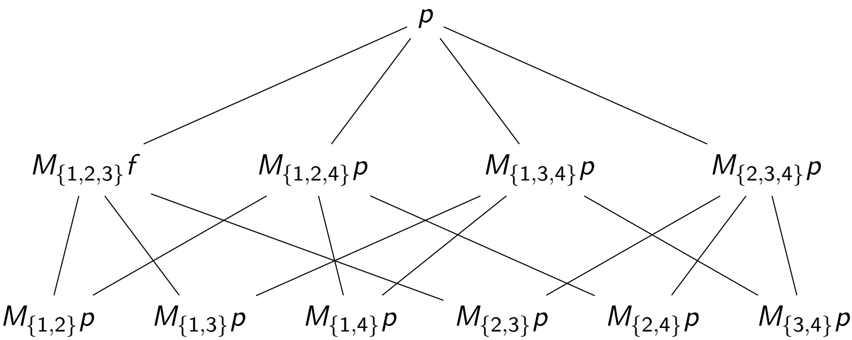
Marginals localize nested levels of rank information

Example

The knowledge of the marginal on $\{a, b, c\}$ induces the knowledge of the marginal on $\{b, c\}$.

$$\begin{aligned}\mathbb{P}[\Sigma(b) < \Sigma(c)] = \ & \mathbb{P}[\Sigma(a) < \Sigma(b) < \Sigma(c)] \\ & + \mathbb{P}[\Sigma(b) < \Sigma(a) < \Sigma(c)] \\ & + \mathbb{P}[\Sigma(b) < \Sigma(c) < \Sigma(a)]\end{aligned}$$

Corresponds to the structure of subsets



$\Rightarrow$ We need to localize the part of information specific to each relative marginal

Another need: The analysis of incomplete rankings

In many situations one only observes **incomplete rankings**

$$a_1 \succ \cdots \succ a_k \qquad \text{with } k \ll n$$

e.g. Users usually express preferences on small subsets of items

Probabilistic modeling

Each observed ranking is modeled as a couple $(\mathbf{A}, \Pi)$ where

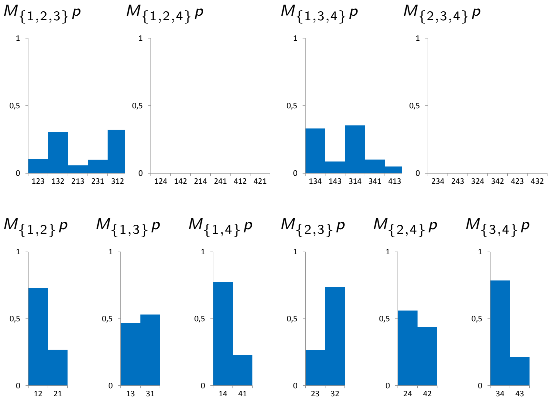
$$\mathbf{A} \sim \nu \qquad \text{probability distribution over } 2^{[\![n]\!]}$$

$$\Pi | (\mathbf{A} = A) \sim M_A p \qquad \text{marginal of } p$$

Another need: The analysis of incomplete rankings

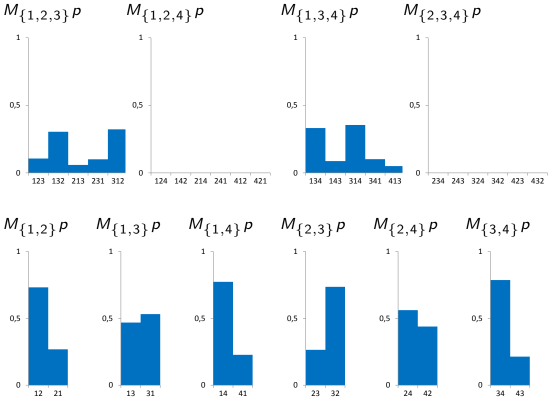
Dataset of incomplete rankings $\mathcal{D}_N = ((\mathbf{A}_1, \Pi_1), \dots, (\mathbf{A}_N, \Pi_N))$

One can construct the empirical estimator of $M_A p$ for A observed



Another need: The analysis of incomplete rankings

- How to combine knowledge inferred on each subset A?
- How can we transfer knowledge between subsets A?



We need to localize information specific to each marginal

Outline

The MRA representation allows to localize the part of information of each relative marginal

Theorem ([Clémençon et al., 2014], informal)

We construct a "wavelet transform"

$$\Psi : p \mapsto (\Psi_B f)_{B \subset [\![n]\!],\ |B| \neq 1}$$

Such that for any $A \subset [\![n]\!]$ *with* $|A| \neq 1$,

$$M_A p \ \text{``}=\text{''}\ \Psi_A p + F((M_B p)_{B \subsetneq A,\ |B| \neq 1})$$

where F *is some function*

The MRA representation allows to localize the part of information of each relative marginal

Example

- $\Psi_{\emptyset} p$ localizes level 0 information

$$\Psi_{\emptyset} p = M_{\emptyset} p := \sum_{\sigma \in \mathfrak{S}_n} p(\sigma)$$

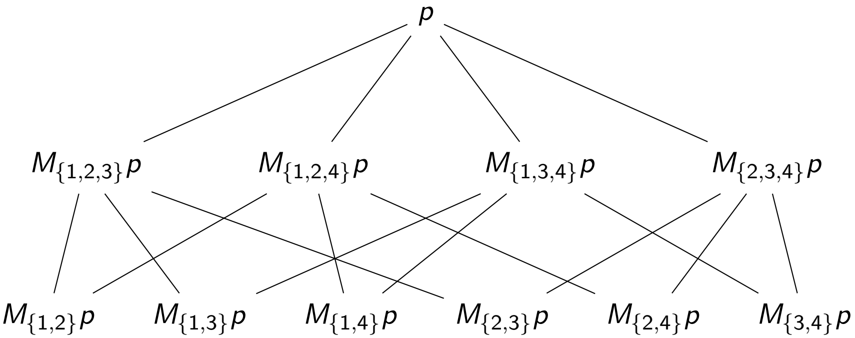
- $\Psi_{\{a,b\}} p$ localizes the part of information specific to $M_{\{a,b\}} p$

$$M_{\{a,b\}} p \text{ ``=''} \; \Psi_{\{a,b\}} p + F(M_{\emptyset} p)$$

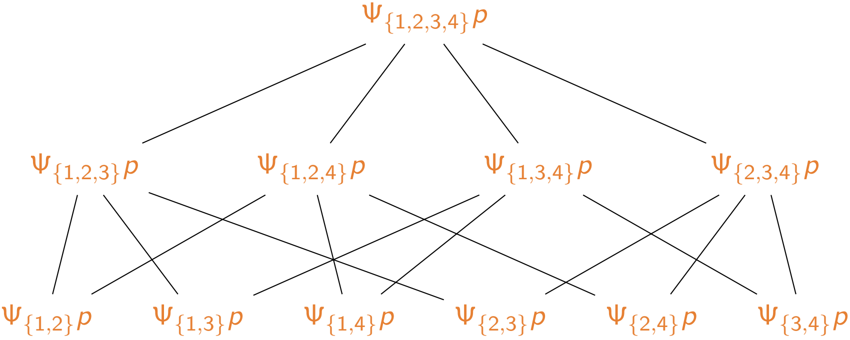
- $\Psi_{\{a,b,c\}} p$ localizes the part of information specific to $M_{\{a,b,c\}} p$

$$M_{\{a,b,c\}} p \text{ ``=''} \; \Psi_{\{a,b,c\}} p + F(M_{\{a,b\}} p + M_{\{a,c\}} p + M_{\{b,c\}} p + M_{\emptyset} p)$$

The MRA representation allows to localize the part of information of each relative marginal



The MRA representation allows to localize the part of information of each relative marginal



Ingredients of the proof

- Linear algebra
- Combinatorics of words
- Recent result in algebraic topology
  (from [Reiner et al., 2013])

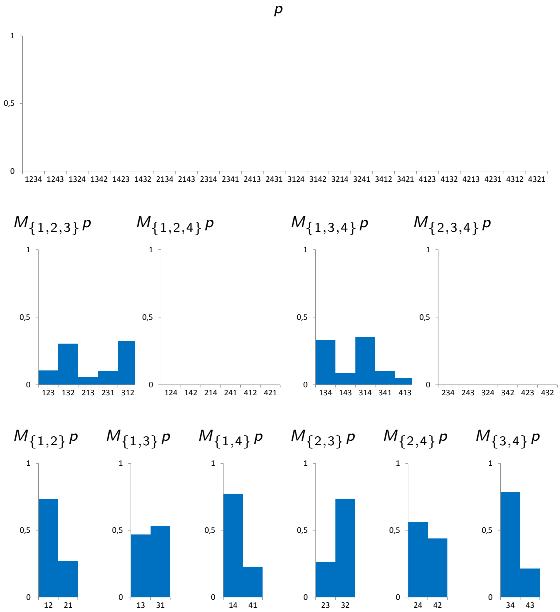
Application 1: solving linear systems

The MRA representation allows to characterize the functions $p : \mathfrak{S}_n \to \mathbb{R}$ with know marginal values:

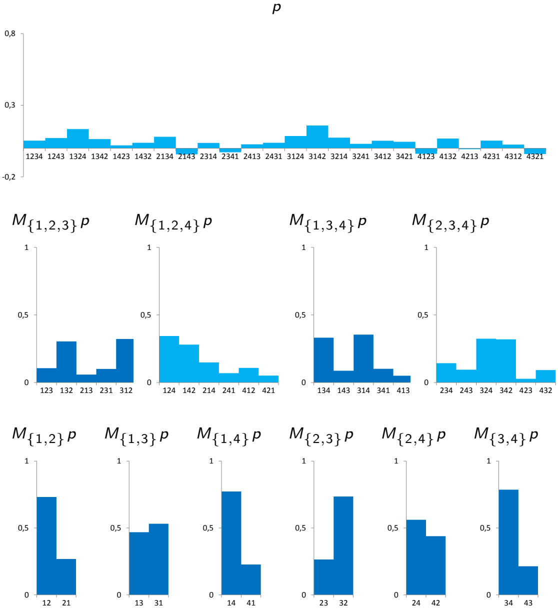
$$M_A p = G_A \qquad \text{for } A \in \mathcal{A}$$

where $\mathcal{A}$ is any collection of subsets.

Application 1: solving linear systems



Application 1: solving linear systems



Application 2: Analysis through relative marginals

Question: which subsets of items capture most of the variability of p?

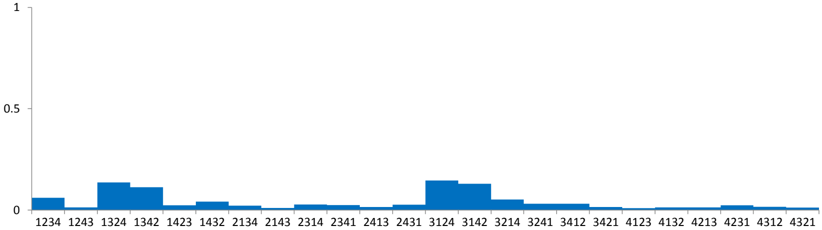
- If we remove the wavelet coefficient $\Psi_A p$ in p then the error of the approximation is

$$(n - |A| + 1)! \|\Psi_A p\|^2$$

NB: the decomposition is not orthogonal, this procedure should be applied as an Orthogonal Matching Pursuit

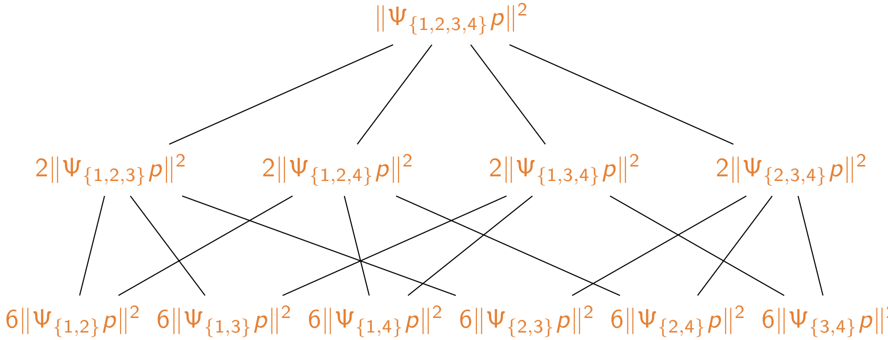
Application 2: Analysis through relative marginals

Example, on the dataset from [Croon, 1989]



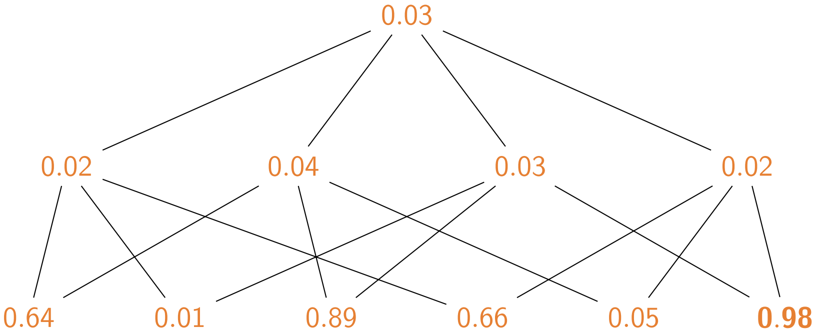
Application 2: Analysis through relative marginals

Example, on the dataset from [Croon, 1989]



Application 2: Analysis through relative marginals

Example, on the dataset from [Croon, 1989]



Application 3: Analysis of incomplete rankings

Dataset of incomplete rankings $\mathcal{D}_N = ((\mathbf{A}_1, \Pi_1), \ldots, (\mathbf{A}_N, \Pi_N))$

General framework

1. Construct empirical estimator $\widehat{P_A}$ for each observed A
2. Compute $\Psi \widehat{P_A}$
3. Compute the global wavelet estimator

$$\widehat{\Psi}_B = \frac{1}{|\{1 \leq i \leq N \mid B \subset \mathbf{A}_i\}|} \sum_A \Psi \widehat{P_A}$$

$\Rightarrow$ Can be computed with a complexity that only depends on the dataset and not on the number of items n

Application 3: Analysis of incomplete rankings

1. Dataset of incomplete rankings
   $\mathcal{D}_N = ((\mathbf{A}_1, \Pi_1), \ldots, (\mathbf{A}_N, \Pi_N))$
2. Compute global wavelet estimator $\widehat{\Psi}_B$
3. Use it to perform a statistical task *in the feature space of wavelet coefficients*

Examples of statistical tasks

- Ranking aggregation
- Regularization for inverse problem
- Conditional prediction
- ...

Outline

Conclusion

Ranking data is fun!

Conclusion

Ranking data is fun!

Its analysis presents great and interesting challenges:

- Most of the maths from euclidean spaces cannot be applied
- But our intuitions still hold

Conclusion

What I did not talk about

- Computational aspects (Fast Wavelet Transform)
- Connection with Fourier analysis
- Many other connections (social choice theory, shuffling, . . .)

What we are working on (future directions)

- Applications to various statistical problems
- How to define efficient *regularization procedures*?
- How to extend to incomplete rankings with ties?
- How to extend to items with features?

Thank you

Clausen, M. and Baum, U. (1993).
Fast fourier transforms for symmetric groups : theory and implementation.
61(204):833–847.

Clémençon, S., Jakubowicz, J., and Sibony, E. (2014).
Multiresolution analysis of incomplete rankings.
ArXiv e-prints.

Croon, M. A. (1989).
Latent class models for the analysis of rankings.
In Geert de Soete, H. F. and Klauer, K. C., editors, *New Developments in Psychological Choice Modeling*, volume 60 of *Advances in Psychology*, pages 99 – 121. North-Holland.

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

Diaconis, P. (1989).
A generalization of spectral analysis with application to ranked data.
The Annals of Statistics, 17(3):949–979.

Huang, J., Guestrin, C., and Guibas, L. (2009).
Fourier theoretic probabilistic inference over permutations.
JMLR, 10:997–1070.

Irurozki, E., Calvo, B., and Lozano, J. (2011).
Learning probability distributions over permutations by means of Fourier coefficients.
Advances in Artificial Intelligence, pages 186–191.

Kakarala, R. (2011).
A signal processing approach to Fourier analysis of ranking data: the importance of phase.
IEEE Transactions on Signal Processing, pages 1–10.

Kondor, R. and Barbosa, M. S. (2010).
Ranking with kernels in Fourier space.

In *Proceedings of COLT'10*, pages 451–463.

Kondor, R. and Dempsey, W. (2012).
Multiresolution analysis on the symmetric group.
In *Neural Information Processing Systems 25*.

Luce, R. D. (1959).
Individual Choice Behavior.
Wiley.

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

Plackett, R. L. (1975).
The analysis of permutations.
Applied Statistics, 2(24):193–202.

Reiner, V., Saliola, F., and Welker, V. (2013).
Spectra of symmetrized shuffling operators.
Memoirs of the American Mathematical Society, 228(1072).

Rockmore, D., Kostelec, P., Hordijk, W., and Stadler, P. F. (2002).
Fast Fourier transform for fitness landscapes.
Applied and Computational Harmonic Analysis, 12(1):57–76.

Sibony, E., Clémençon, S., and Jakubowicz, J. (2015).
Mra-based statistical learning from incomplete rankings.
In *Proceedings of the 32nd International Conference on Machine Learning, ICML 2015, Lille, France, 6-11 July 2015*, pages 1432–1441.

Thurstone, L. L. (1927).
A law of comparative judgment.
Psychological Review, 34(4):273–286.