Riffled Independence for Ranked Data

Jonathan Huang and Carlos Guestrin
NIPS, 2009

Zachary Schutzman

University of Pennsylvania

February 2, 2017

Presentation Outline

High-Level Idea

High-Level Idea

Distributions over permutations show up in a lot of contexts

High-Level Idea

Distributions over permutations show up in a lot of contexts

- Object tracking
- Feature mapping
- **RANKED DATA!**

High-Level Idea

Distributions over permutations show up in a lot of contexts

- Object tracking
- Feature mapping
- **RANKED DATA!**

Are there any mathematical techniques that can tease out useful information from distributions?

High-Level Idea

Distributions over permutations show up in a lot of contexts

- Object tracking
- Feature mapping
- **RANKED DATA!**

Are there any mathematical techniques that can tease out useful information from distributions?

- Yes!

Notation!

Notation!

S_n is the symmetric group on n elements
(the set of bijections $\pi : [n] \to [n]$)

Notation!

S_n is the symmetric group on n elements
(the set of bijections $\pi : [n] \to [n]$)

A distribution over S_n is a function $h : S_n \to \mathbb{R}$ such that:

- for all $\sigma \in S_n$, $h(\sigma) \geq 0$
- $\sum_{\sigma \in S_n} h(\sigma) = 1$

Full Independence

Let h be a distribution over S_n and let σ_P be a p-subset of $[n]$, and σ_Q its complement

Full Independence

Let h be a distribution over S_n and let σ_P be a p-subset of $[n]$, and σ_Q its complement

h is **probabilistically independent** if there exist distributions f and g over S_p and S_q such that
$h(\sigma) = f(\sigma_P) \cdot g(\sigma_Q)$

Full Independence

Let h be a distribution over S_n and let σ_P be a p-subset of $[n]$, and σ_Q its complement

h is **probabilistically independent** if there exist distributions f and g over S_p and S_q such that
$h(\sigma) = f(\sigma_P) \cdot g(\sigma_Q)$

Why is this nice?

Full Independence

Let h be a distribution over S_n and let σ_P be a p-subset of $[n]$, and σ_Q its complement

h is **probabilistically independent** if there exist distributions f and g over S_p and S_q such that
$h(\sigma) = f(\sigma_P) \cdot g(\sigma_Q)$

Why is this nice?
If h is probabilistically independent, it only has $p! + q!$ parameters

Full Independence

Let h be a distribution over S_n and let σ_P be a p-subset of $[n]$, and σ_Q its complement

h is **probabilistically independent** if there exist distributions f and g over S_p and S_q such that
$h(\sigma) = f(\sigma_P) \cdot g(\sigma_Q)$

Why is this nice?
If h is probabilistically independent, it only has $p! + q!$ parameters

But!

Full Independence

Let h be a distribution over S_n and let σ_P be a p-subset of $[n]$, and σ_Q its complement

h is **probabilistically independent** if there exist distributions f and g over S_p and S_q such that
$h(\sigma) = f(\sigma_P) \cdot g(\sigma_Q)$

Why is this nice?
If h is probabilistically independent, it only has $p! + q!$ parameters

But!
This is a *very* strong assumption to make for ranked data

Riffle Independence

Let's define a looser form of independence

Let's define a looser form of independence

A (p, q)-interleaving is an element of S_n which preserves the ordering within the p- and q-subsets. Call $\Omega_{p,q} \subset S_n$ the set of such permutations

Riffle Independence

Let's define a looser form of independence

A (p,q)-interleaving is an element of S_n which preserves the ordering within the p- and q-subsets. Call $\Omega_{p,q} \subset S_n$ the set of such permutations

A **riffle distribution** over Σ_n is a distribution which assigns non-zero probability only to elements of $\Omega_{p,q}$

Riffle Independence

Let's define a looser form of independence

A (p, q)-interleaving is an element of S_n which preserves the ordering within the p- and q-subsets. Call $\Omega_{p,q} \subset S_n$ the set of such permutations

A **riffle distribution** over Σ_n is a distribution which assigns non-zero probability only to elements of $\Omega_{p,q}$

h is **riffle independent** if there exist distributions f and g and a riffle distribution m such that
$h(\sigma) = m * (f(\sigma_P) \cdot g(\sigma_Q))$

Riffle Shuffles

m can be an arbitrary distribution on $\Omega_{p,q}$, but we might find it useful to think about some restrictions

Riffle Shuffles

m can be an arbitrary distribution on $\Omega_{p,q}$, but we might find it useful to think about some restrictions

A **uniform** riffle distribution m^{unif} assigns probability $\frac{1}{|\Omega_{p,q}|}$ to each element of $\Omega_{p,q}$

Riffle Shuffles

m can be an arbitrary distribution on $\Omega_{p,q}$, but we might find it useful to think about some restrictions

A **uniform** riffle distribution m^{unif} assigns probability $\frac{1}{|\Omega_{p,q}|}$ to each element of $\Omega_{p,q}$

A **biased** riffle distribution with parameter α as an interleaving with an α-weighted preference for items in group p over those in group q

Riffle Shuffles

m can be an arbitrary distribution on $\Omega_{p,q}$, but we might find it useful to think about some restrictions

A **uniform** riffle distribution m^{unif} assigns probability $\frac{1}{|\Omega_{p,q}|}$ to each element of $\Omega_{p,q}$

A **biased** riffle distribution with parameter α as an interleaving with an α-weighted preference for items in group p over those in group q

- if $\alpha = .5$, we get the uniform distribution
- if $\alpha = 1$ we 'interleave' by putting all the items in p first, then all the items in q $(m(123 \ldots n) = 1)$
- if $\alpha = 0$, we put all the items in q first

Riffle Shuffles

h is **riffle independent** if there exist distributions f and g and a riffle distribution m such that
$h(\sigma) = m * (f(\sigma_P) \cdot g(\sigma_Q))$

Riffle Shuffles

h is **riffle independent** if there exist distributions f and g and a riffle distribution m such that
$h(\sigma) = m * (f(\sigma_P) \cdot g(\sigma_Q))$

The authors present **Fourier-domain algorithms** to analyze and interpret riffle distributions

Presentation Outline

Groups

What is a group?

Groups

What is a group?

- A group G is a set and an operation $\circ$ satisfying the following properties:

Groups

What is a group?

- A group G is a set and an operation $\circ$ satisfying the following properties:
  - Closure: for all $g, h \in G$, $g \circ h \in G$

Groups

What is a group?

- A group G is a set and an operation $\circ$ satisfying the following properties:
  - Closure: for all $g, h \in G$, $g \circ h \in G$
  - Identity: there is a unique $e \in G$ such that for all $g \in G$, $e \circ g = g \circ e = g$

Groups

What is a group?

- A group G is a set and an operation $\circ$ satisfying the following properties:
  - Closure: for all $g, h \in G$, $g \circ h \in G$
  - Identity: there is a unique $e \in G$ such that for all $g \in G$, $e \circ g = g \circ e = g$
  - Inverses: for each $g \in G$, there is a unique g^{-1} such that $g^{-1} \circ g = g \circ g^{-1} = e$

Groups

What is a group?

- A group G is a set and an operation $\circ$ satisfying the following properties:
  - Closure: for all $g, h \in G$, $g \circ h \in G$
  - Identity: there is a unique $e \in G$ such that for all $g \in G$, $e \circ g = g \circ e = g$
  - Inverses: for each $g \in G$, there is a unique g^{-1} such that $g^{-1} \circ g = g \circ g^{-1} = e$
  - Associativity: for all $f, g, h \in G$, $f \circ (g \circ h) = (f \circ g) \circ h$

The Symmetric Group

S_n forms a group with respect to the operation of composition of permutations

The Symmetric Group

S_n forms a group with respect to the operation of composition of permutations

We can write group elements in cycle notation, which describes a way of shifting the objects to form a permutation.

The Symmetric Group

S_n forms a group with respect to the operation of composition of permutations

We can write group elements in cycle notation, which describes a way of shifting the objects to form a permutation.

- **Example:** The permutation 15324

The Symmetric Group

S_n forms a group with respect to the operation of composition of permutations

We can write group elements in cycle notation, which describes a way of shifting the objects to form a permutation.

- **Example:** The permutation 15324 can be written as (1)

The Symmetric Group

S_n forms a group with respect to the operation of composition of permutations

We can write group elements in cycle notation, which describes a way of shifting the objects to form a permutation.

- **Example:** The permutation 15324 can be written as (1)(2

The Symmetric Group

S_n forms a group with respect to the operation of composition of permutations

We can write group elements in cycle notation, which describes a way of shifting the objects to form a permutation.

- **Example:** The permutation 15324 can be written as (1)(24

The Symmetric Group

S_n forms a group with respect to the operation of composition of permutations

We can write group elements in cycle notation, which describes a way of shifting the objects to form a permutation.

- **Example:** The permutation 15324 can be written as (1)(245

The Symmetric Group

S_n forms a group with respect to the operation of composition of permutations

We can write group elements in cycle notation, which describes a way of shifting the objects to form a permutation.

- **Example:** The permutation 15324 can be written as (1)(245)

The Symmetric Group

S_n forms a group with respect to the operation of composition of permutations

We can write group elements in cycle notation, which describes a way of shifting the objects to form a permutation.

- **Example:** The permutation 15324 can be written as (1)(245)(3)

The Symmetric Group

S_n forms a group with respect to the operation of composition of permutations

We can write group elements in cycle notation, which describes a way of shifting the objects to form a permutation.

- **Example:** The permutation 15324 can be written as (1)(245)(3)
- The identity element is $(1)(2)\ldots(n)$

Group Representations

It is useful to be able to think about group elements as matrices, rather than objects in a set.

Group Representations

It is useful to be able to think about group elements as matrices, rather than objects in a set. We know how to do a lot of math with matrices

Group Representations

It is useful to be able to think about group elements as matrices, rather than objects in a set. We know how to do a lot of math with matrices

A **representation** of a group is a function $\rho : G \to \mathbb{C}^{k \times k}$ such that for all $g, h \in G$, $\rho(gh) = \rho(g)\rho(h)$

Group Representations

It is useful to be able to think about group elements as matrices, rather than objects in a set. We know how to do a lot of math with matrices

A **representation** of a group is a function $\rho : G \to \mathbb{C}^{k \times k}$ such that for all $g, h \in G$, $\rho(gh) = \rho(g)\rho(h)$

A representation is **irreducible** if it cannot be written as the direct sum of two other representations

Representations of S_3 (a worked example)

Let's think about the group S_3. There are only six elements and three irreducible representations

Let's think about the group S_3. There are only six elements and three irreducible representations

π
123
213
132
321
231
312

Representations of S_3 (a worked example)

Let's think about the group S_3. There are only six elements and three irreducible representations

π	g
123	()
213	(12)
132	(23)
321	(13)
231	(132)
312	(123)

Representations of S_3 (a worked example)

Let's think about the group S_3. There are only six elements and three irreducible representations

π	g	ρ_{triv}
123	()	1
213	(12)	1
132	(23)	1
321	(13)	1
231	(132)	1
312	(123)	1

Representations of S_3 (a worked example)

Let's think about the group S_3. There are only six elements and three irreducible representations

π	g	ρ_{triv}	ρ_{sgn}
123	$()$	1	1
213	(12)	1	-1
132	(23)	1	-1
321	(13)	1	-1
231	(132)	1	1
312	(123)	1	1

Representations of S_3 (a worked example)

Let's think about the group S_3. There are only six elements and three irreducible representations

π	g	ρ_{triv}	ρ_{sgn}	ρ_{std}
123	$()$	1	1	$\begin{pmatrix}1 & 0\\0 & 1\end{pmatrix}$
213	(12)	1	-1	$\begin{pmatrix}-1 & 1\\0 & 1\end{pmatrix}$
132	(23)	1	-1	$\begin{pmatrix}1 & 0\\1 & -1\end{pmatrix}$
321	(13)	1	-1	$\begin{pmatrix}0 & -1\\-1 & 0\end{pmatrix}$
231	(132)	1	1	$\begin{pmatrix}-1 & 1\\-1 & 0\end{pmatrix}$
312	(123)	1	1	$\begin{pmatrix}0 & -1\\1 & -1\end{pmatrix}$

The Fourier transform

We can define the **Fourier transform** of a function at a representation as follows:

The Fourier transform

We can define the **Fourier transform** of a function at a representation as follows:

Let $f(g)$ be a function mapping group elements to real numbers and let ρ be any representation of G

The Fourier transform

We can define the **Fourier transform** of a function at a representation as follows:

Let $f(g)$ be a function mapping group elements to real numbers and let ρ be any representation of G

Then the Fourier transform of f at ρ is equal to $\hat{f}_\rho = \sum_{g \in G} f(g)\rho(g)$

The Fourier transform

We can define the **Fourier transform** of a function at a representation as follows:

Let $f(g)$ be a function mapping group elements to real numbers and let ρ be any representation of G

Then the Fourier transform of f at ρ is equal to $\hat{f}_\rho = \sum_{g \in G} f(g)\rho(g)$

If f is a distribution (h from before), then $\hat{h}$ at the irreducibles systematically encodes h.

A simple worked example:

$$\begin{array}{c} \pi \\ () \\ (12) \\ (23) \\ (13) \\ (132) \\ (123) \end{array}$$

A simple worked example:

$$\hat{h}_{\rho_{triv}} = \frac{2}{12}\left[1\right] + \frac{2}{12}\left[1\right] + \frac{3}{12}\left[1\right] + \frac{1}{12}\left[1\right] + \frac{1}{12}\left[1\right] + \frac{3}{12}\left[1\right] = \left[1\right]$$

π	$h(\pi)$
()	$2/12$
(12)	$2/12$
(23)	$3/12$
(13)	$1/12$
(132)	$1/12$
(123)	$3/12$

A simple worked example:

$$\hat{h}_{\rho_{triv}} = \frac{2}{12}\begin{bmatrix}1\end{bmatrix} + \frac{2}{12}\begin{bmatrix}1\end{bmatrix} + \frac{3}{12}\begin{bmatrix}1\end{bmatrix} + \frac{1}{12}\begin{bmatrix}1\end{bmatrix} + \frac{1}{12}\begin{bmatrix}1\end{bmatrix} + \frac{3}{12}\begin{bmatrix}1\end{bmatrix} = \begin{bmatrix}1\end{bmatrix}$$

π	$h(\pi)$
()	$2/12$
(12)	$2/12$
(23)	$3/12$
(13)	$1/12$
(132)	$1/12$
(123)	$3/12$

$$\hat{h}_{\rho_{std}} = \frac{2}{12}\begin{pmatrix}1 & 0\\ 0 & 1\end{pmatrix} + \frac{2}{12}\begin{pmatrix}-1 & 1\\ 0 & 1\end{pmatrix} + \frac{3}{12}\begin{pmatrix}1 & 0\\ 1 & -1\end{pmatrix} + \frac{1}{12}\begin{pmatrix}0 & -1\\ -1 & 0\end{pmatrix} + \frac{1}{12}\begin{pmatrix}-1 & 1\\ -1 & 0\end{pmatrix} + \frac{3}{12}\begin{pmatrix}0 & -1\\ 1 & -1\end{pmatrix} = \frac{1}{12}\begin{pmatrix}2 & -1\\ 4 & -2\end{pmatrix}$$

A simple worked example:

$$\hat{h}_{\rho_{triv}} = \frac{2}{12}\left[1\right] + \frac{2}{12}\left[1\right] + \frac{3}{12}\left[1\right] + \frac{1}{12}\left[1\right] + \frac{1}{12}\left[1\right] + \frac{3}{12}\left[1\right] = \left[1\right]$$

π	$h(\pi)$
$()$	$2/12$
(12)	$2/12$
(23)	$3/12$
(13)	$1/12$
(132)	$1/12$
(123)	$3/12$

$$\hat{h}_{\rho_{std}} = \frac{2}{12}\begin{pmatrix}1 & 0\\0 & 1\end{pmatrix} + \frac{2}{12}\begin{pmatrix}-1 & 1\\0 & 1\end{pmatrix} + \frac{3}{12}\begin{pmatrix}1 & 0\\1 & -1\end{pmatrix} + \frac{1}{12}\begin{pmatrix}0 & -1\\-1 & 0\end{pmatrix} + \frac{1}{12}\begin{pmatrix}-1 & 1\\-1 & 0\end{pmatrix} + \frac{3}{12}\begin{pmatrix}0 & -1\\1 & -1\end{pmatrix} = \frac{1}{12}\begin{pmatrix}2 & -1\\4 & -2\end{pmatrix}$$

$$\hat{h}_{\rho_{sgn}} = \frac{2}{12}\left[1\right] + \frac{2}{12}\left[-1\right] + \frac{3}{12}\left[-1\right] + \frac{1}{12}\left[1\right] + \frac{1}{12}\left[1\right] + \frac{3}{12}\left[-1\right] = \frac{1}{12}\left[0\right]$$

Example, continued

We have our three Fourier matrices:

Example, continued

We have our three Fourier matrices:

$\hat{h}_{\rho_{triv}} = \begin{bmatrix} 1 \end{bmatrix}$

Example, continued

We have our three Fourier matrices:

$$\hat{h}_{\rho_{triv}} = \begin{bmatrix} 1 \end{bmatrix}$$

$$\hat{h}_{\rho_{std}} = \frac{1}{12}\begin{pmatrix} 2 & -1 \\ 4 & -2 \end{pmatrix}$$

Example, continued

We have our three Fourier matrices:

$$\hat{h}_{\rho_{triv}} = \begin{bmatrix}1\end{bmatrix}$$
$$\hat{h}_{\rho_{std}} = \frac{1}{12}\begin{pmatrix}2 & -1 \\ 4 & -2\end{pmatrix}$$
$$\hat{h}_{\rho_{sgn}} = \begin{bmatrix}0\end{bmatrix}$$

Example, continued

We have our three Fourier matrices:

$$\hat{h}_{\rho_{triv}} = \begin{bmatrix} 1 \end{bmatrix}$$
$$\hat{h}_{\rho_{std}} = \frac{1}{12}\begin{pmatrix} 2 & -1 \\ 4 & -2 \end{pmatrix}$$
$$\hat{h}_{\rho_{sgn}} = \begin{bmatrix} 0 \end{bmatrix}$$

Let's glue them together

Example, continued

We have our three Fourier matrices:

$$\hat{h}_{\rho_{triv}} = \begin{bmatrix} 1 \end{bmatrix}$$
$$\hat{h}_{\rho_{std}} = \frac{1}{12}\begin{pmatrix} 2 & -1 \\ 4 & -2 \end{pmatrix}$$
$$\hat{h}_{\rho_{sgn}} = \begin{bmatrix} 0 \end{bmatrix}$$

Let's glue them together

$$\frac{1}{12}\begin{pmatrix} \mathbf{0} & 0 & 0 & 0 \\ 0 & \mathbf{2} & \mathbf{-1} & 0 \\ 0 & \mathbf{4} & \mathbf{-2} & 0 \\ 0 & 0 & 0 & \mathbf{1} \end{pmatrix}$$

Example, continued

We have our three Fourier matrices:

$$\hat{h}_{\rho_{triv}} = \begin{bmatrix} 1 \end{bmatrix}$$
$$\hat{h}_{\rho_{std}} = \frac{1}{12}\begin{pmatrix} 2 & -1 \\ 4 & -2 \end{pmatrix}$$
$$\hat{h}_{\rho_{sgn}} = \begin{bmatrix} 0 \end{bmatrix}$$

Let's glue them together and change the basis

$$\frac{1}{12}\begin{pmatrix} \mathbf{0} & 0 & 0 & 0 \\ 0 & \mathbf{2} & \mathbf{-1} & 0 \\ 0 & \mathbf{4} & \mathbf{-2} & 0 \\ 0 & 0 & 0 & \mathbf{1} \end{pmatrix}$$

Example, continued

We have our three Fourier matrices:

$$\hat{h}_{\rho_{triv}} = \begin{bmatrix} 1 \end{bmatrix}$$
$$\hat{h}_{\rho_{std}} = \frac{1}{12}\begin{pmatrix} 2 & -1 \\ 4 & -2 \end{pmatrix}$$
$$\hat{h}_{\rho_{sgn}} = \begin{bmatrix} 0 \end{bmatrix}$$

Let's glue them together and change the basis

$$\begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1 \\ 0 & -1 & 1 & 1 \\ 0 & 0 & -1 & 1 \end{pmatrix} \frac{1}{12} \begin{pmatrix} \mathbf{0} & 0 & 0 & 0 \\ 0 & \mathbf{2} & \mathbf{-1} & 0 \\ 0 & \mathbf{4} & \mathbf{-2} & 0 \\ 0 & 0 & 0 & \mathbf{1} \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1 \\ 0 & -1 & 1 & 1 \\ 0 & 0 & -1 & 1 \end{pmatrix}^{-1}$$

Example, continued

We have our three Fourier matrices:

$$\hat{h}_{\rho_{triv}} = \begin{bmatrix} 1 \end{bmatrix}$$
$$\hat{h}_{\rho_{std}} = \frac{1}{12}\begin{pmatrix} 2 & -1 \\ 4 & -2 \end{pmatrix}$$
$$\hat{h}_{\rho_{sgn}} = \begin{bmatrix} 0 \end{bmatrix}$$

Let's glue them together and change the basis

$$\begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1 \\ 0 & -1 & 1 & 1 \\ 0 & 0 & -1 & 1 \end{pmatrix} \frac{1}{12} \begin{pmatrix} \mathbf{0} & 0 & 0 & 0 \\ 0 & \mathbf{2} & \mathbf{-1} & 0 \\ 0 & \mathbf{4} & \mathbf{-2} & 0 \\ 0 & 0 & 0 & \mathbf{1} \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1 \\ 0 & -1 & 1 & 1 \\ 0 & 0 & -1 & 1 \end{pmatrix}^{-1} = \frac{1}{12} \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 5 & 3 & 4 \\ 0 & 5 & 3 & 4 \\ 0 & 2 & 6 & 4 \end{pmatrix}$$

Example, interpreted:

Distribution:

π	$h(\pi)$
$()$	$2/12$
(12)	$2/12$
(23)	$3/12$
(13)	$1/12$
(132)	$1/12$
(123)	$3/12$

Example, interpreted:

Distribution:

π	$h(\pi)$
$()$	$2/12$
(12)	$2/12$
(23)	$3/12$
(13)	$1/12$
(132)	$1/12$
(123)	$3/12$

Recall we had $\hat{h}_{\rho_{triv}} = (1)$. This confirms the values of h sum to 1.

Example, interpreted:

Distribution:

π	$h(\pi)$
$()$	$2/12$
(12)	$2/12$
(23)	$3/12$
(13)	$1/12$
(132)	$1/12$
(123)	$3/12$

Recall we had $\hat{h}_{\rho_{triv}} = \begin{pmatrix}1\end{pmatrix}$. This confirms the values of h sum to 1.

Our transformed matrix is

$$\frac{1}{12}\begin{pmatrix}0 & 0 & 0 & 0\\ 0 & 5 & 3 & 4\\ 0 & 5 & 3 & 4\\ 0 & 2 & 6 & 4\end{pmatrix}$$

Presentation Outline

Back to our regularly scheduled paper

Let's quickly remind ourselves:

Back to our regularly scheduled paper

Let's quickly remind ourselves:

Fully independent: $h = f \cdot g$
Riffle independent: $h = m * (f \cdot g)$
$\Omega_{p,q}$ is the set of interleavings

Back to our regularly scheduled paper

Let's quickly remind ourselves:

Fully independent: $h = f \cdot g$
Riffle independent: $h = m * (f \cdot g)$
$\Omega_{p,q}$ is the set of interleavings

Riffle independence is a generalization of full independence.
Indeed, if m is a delta distribution, we recover full independence.

Biased riffle shuffles

Let's think about m^{unif} as the uniform distribution over $\Omega_{p,q}$. This in a sense reflects preferences within groups, but indifference across groups.

Biased riffle shuffles

Let's think about m^{unif} as the uniform distribution over $\Omega_{p,q}$. This in a sense reflects preferences within groups, but indifference across groups.

What if we want to think about preferences across groups?

Biased riffle shuffles

Let's think about m^{unif} as the uniform distribution over $\Omega_{p,q}$. This in a sense reflects preferences within groups, but indifference across groups.

What if we want to think about preferences across groups?

The authors introduce the idea of a **biased riffle shuffle**, with bias parameter α.

Biased riffle shuffles

Let's think about m^{unif} as the uniform distribution over $\Omega_{p,q}$. This in a sense reflects preferences within groups, but indifference across groups.

What if we want to think about preferences across groups?

The authors introduce the idea of a **biased riffle shuffle**, with bias parameter α.
This weights objects in the p-subset proportional to αp.

Between conditional and full independence

Let's recall that to specify a full distribution over S_n we need $n!$ parameters.

Between conditional and full independence

Let's recall that to specify a full distribution over S_n we need $n!$ parameters.

Full independence requires only $p! + q!$.

Between conditional and full independence

Let's recall that to specify a full distribution over S_n we need $n!$ parameters.

Full independence requires only $p! + q!$.

So how about conditional independence? Is riffle independence the same?

Between conditional and full independence

Let's recall that to specify a full distribution over S_n we need $n!$ parameters.

Full independence requires only $p! + q!$.

So how about conditional independence? Is riffle independence the same?

No! To see this, notice that conditional independence, where we pick a subset and condition f and g on that subset has $\binom{n}{p} \times (p! + q!)$ parameters while riffle independence requires $\binom{n}{p} + p! + q!$

Between conditional and full independence

Let's recall that to specify a full distribution over S_n we need $n!$ parameters.

Full independence requires only $p! + q!$.

So how about conditional independence? Is riffle independence the same?

No! To see this, notice that conditional independence, where we pick a subset and condition f and g on that subset has $\binom{n}{p} \times (p! + q!)$ parameters while riffle independence requires $\binom{n}{p} + p! + q!$

Nevertheless, we can do things like MAP estimation and probabilistic decomposition over the f and g riffle factors as if we had conditional independence.

Presentation Outline

RiffleJoin and RiffleSplit

The authors define two algorithms to combine riffle factors f, g, m into an h and to split an h into f, g.

RiffleJoin and RiffleSplit

The authors define two algorithms to combine riffle factors f, g, m into an h and to split an h into f, g.

These algorithms are really complicated and (I think) unintuitive without a lot of knowledge of Fourier analysis.
See citations [6],[7],[8] for the math.

RiffleJoin and RiffleSplit

The authors define two algorithms to combine riffle factors f, g, m into an h and to split an h into f, g.

These algorithms are really complicated and (I think) unintuitive without a lot of knowledge of Fourier analysis.
See citations [6],[7],[8] for the math.

I hope that the example from before at least convinces you that such a decomposition may be possible in the Fourier domain.

Some experiments

The authors tested their algorithm on two data sets: the American Psychological Association election and the Sushi Data Set.

Experiment 1: APA Election

The data: 5738 ballots of ranked preferences across 5 candidates for president of the APA

Experiment 1: APA Election

The data: 5738 ballots of ranked preferences across 5 candidates for president of the APA

The hypothesis: C1,C3 and C4,C5 fall on opposite ends of a political spectrum. Voters may express a preference for one group over the other, then express preferences within the group.

Experiment 1: APA Election

The data: 5738 ballots of ranked preferences across 5 candidates for president of the APA
The hypothesis: C1,C3 and C4,C5 fall on opposite ends of a political spectrum. Voters may express a preference for one group over the other, then express preferences within the group.
The experiment: ignoring C2, decompose the full distribution into riffle factors.

Experiment 1: APA Election

The data: 5738 ballots of ranked preferences across 5 candidates for president of the APA
The hypothesis: C1,C3 and C4,C5 fall on opposite ends of a political spectrum. Voters may express a preference for one group over the other, then express preferences within the group.
The experiment: ignoring C2, decompose the full distribution into riffle factors.

The results: The authors find that C1,C3 and C4,C5 are nearly riffle independent, and that fitting a mixture-of-riffles model yields bias parameters that strongly suggest their hypothesis is correct.

Experiement 2: Sushi

The data: 5000 full rankings of 10 different kinds of sushi (note, WAY fewer than 10! observations). The authors divide this into two groups of five.

Experiement 2: Sushi

The data: 5000 full rankings of 10 different kinds of sushi (note, WAY fewer than 10! observations). The authors divide this into two groups of five.

The hypothesis: the full distribution can be approximated by combining an optimal *m* with two factor subsets and/or by finding an optimal bias parameter α.

Experiement 2: Sushi

The data: 5000 full rankings of 10 different kinds of sushi (note, WAY fewer than 10! observations). The authors divide this into two groups of five.

The hypothesis: the full distribution can be approximated by combining an optimal *m* with two factor subsets and/or by finding an optimal bias parameter α.

The experiment: Try to learn the true distribution using these two methods

Experiement 2: Sushi

The data: 5000 full rankings of 10 different kinds of sushi (note, WAY fewer than 10! observations). The authors divide this into two groups of five.
The hypothesis: the full distribution can be approximated by combining an optimal m with two factor subsets and/or by finding an optimal bias parameter α.
The experiment: Try to learn the true distribution using these two methods

The results: Assuming riffle independence significantly lowers the sample complexity required to learn the distribution.

Experiement 2: Sushi

The data: 5000 full rankings of 10 different kinds of sushi (note, WAY fewer than 10! observations). The authors divide this into two groups of five.
The hypothesis: the full distribution can be approximated by combining an optimal *m* with two factor subsets and/or by finding an optimal bias parameter α.
The experiment: Try to learn the true distribution using these two methods

The results: Assuming riffle independence significantly lowers the sample complexity required to learn the distribution.
Biased riffle shuffles are a useful tool for learning on small samples.

Presentation Outline

Discussion

What are some problems where riffle independence might lend some new insight?

What if the riffle factors are non-obvious?

How about if there are 3+ subsets we want to study?

Anything else?