Induced Probability Distributions on Double Cosets

Mackenzie Simper
Joint work with Persi Diaconis
Stanford University

MIT
April 8, 2021
Outline

Part 1 –

1. Background on double cosets

2. Questions: What do double cosets look like? What distribution is induced by uniform distribution?

3. Three examples:
   1. Mallows measure on permutations
   2. Ewens sampling formula on partitions
   3. Fisher-Yates distribution on contingency tables

Part 2 – Random transpositions Markov chain on contingency tables
Double cosets

Let $G$ be a finite group, $H$ and $K$ subgroups. Define an equivalence relation on $G$ by

$$s \sim t \iff s = h^{-1}tk \text{ for } s, t \in G, h \in H, k \in K.$$ 

**Definition** The double-cosets $H \backslash G / K$ is the set of equivalence classes. The double coset containing $s \in G$ is written $HsK$. That is,

$$HsK = \{ hsk : h \in H, k \in K \}.$$ 

Think: Elements in a double coset are equivalent under the symmetries coming from two subgroups.
Simple example

• $G = S_4$

• $H = K = \{1234, 1324, 2134, 2314, 3124, 3214\} = S_3 \subset S_4$. That is, $H$ is the set of all permutations $\sigma \in S_4$ that fix 4 ($\sigma(4) = 4$).

• There are two double cosets:

  1. Double coset containing the identity –
     $H(id)H = \{1234, 1324, 2134, 2314, 3124, 3214\}$

  2. $\sigma = 4123$, then $H\sigma H$ is the set of all permutations which do not fix 4.
Questions

• How many double cosets are there?

• What do the double cosets ‘look like’?

• Pick an element $g \in G$ uniformly at random. What double coset is it likely to be in? In other words: what distribution on double cosets is induced by the uniform distribution on $G$?

See our paper ‘Statistical Enumeration of Groups by Double Cosets’ on arxiv for many more details.
Simple example

• $G = S_4$, 
  $H = K = \{1234, 1324, 2134, 2314, 3124, 3214\} = S_3 \subset S_4$.

• There are two double cosets:

  1. Double coset containing the identity – 
     $H(id)H = \{1234, 1324, 2134, 2314, 3124, 3214\}$

  2. $\sigma = 4123$, then $H\sigma H$ is the set of all permutations which do not fix 4.

• Induced distribution:

  1. $p(H(id)H) = \frac{|H(id)H|}{|G|} = \frac{3!}{4!} = \frac{1}{4}$

  2. $p(H\sigma H) = \frac{|H\sigma H|}{|G|} = \frac{3}{4}$
Example 1 – Mallows measure on permutations

- $G = GL_n(\mathbb{F}_q)$, the general linear group of a field with $q$ elements. $H = K = B$, the lower triangular matrices in $G$.

- If $W \cong S_n$ is the permutation group embedded in $G$ as permutation matrices, then (Bruhat decomposition)

$$G = \bigcup_{\omega \in W} B\omega B$$

- The induced measure on $S_n$ is the Mallows measure

$$p_q(\omega) = \frac{q^{I(\omega)}}{[n]}$$

where $I(\omega)$ is the number of inversions in the permutation $\omega$ and $[n] = (1+q)(1+q+q^2)\ldots(1+q+\ldots+q^{n-1})$. 
Example 1 – Mallows measure on permutations

- $G = GL_n(\mathbb{F}_q)$, the general linear group of a field with $q$ elements. $H = K = B$, the lower triangular matrices in $G$.

- If $W \cong S_n$ is the permutation group embedded in $G$ as permutation matrices, then (Bruhat decomposition)

  $$G = \bigcup_{\omega \in W} B\omega B$$

- Thus, double cosets are indexed by permutations. The induced measure on $S_n$ is the Mallows measure

  $$\rho_q(\omega) = \frac{q^{l(\omega)}}{[n]_q!},$$

  where $l(\omega)$ is the number of inversions in the permutation $\omega$ and $[n]_q! = (1 + q)(1 + q + q^2)\ldots(1 + q + \ldots + q^{n-1})$. 
Where does this come from?

- Lower triangular matrices $|B| = (q - 1)^n \cdot q^{\binom{n}{2}}$
- $|G| = |B| \prod_{i=1}^{n-1} (1 + q + \ldots + q^i) = |B| \cdot [n]_q$
- $|B\omega B| = |B|q^{I(\omega)}$, where $I(\omega)$ is number of inversions
- Combining these identities, can see the induced distribution on double cosets:

$$p_q(\omega) = \frac{|B\omega B|}{|G|} = \frac{q^{I(\omega)}}{[n]_q}.$$
Example 2 – Ewens measure on partitions

• \( G = S_{2n}, \ H = K = B_n \) the group of symmetries of a \( n \)-dimensional hypercube; \( B_n \subset S_{2n} \) as the subgroup of centrally symmetric permutations, i.e.
  \[ \sigma(i) + \sigma(2n + 1 - i) = 2n + 1. \]

• Example: \( B_2 = \{1234, 4231, 1324, 4321, 3142, 2143, 3412, 2413\} \subset S_4 \)

• \( |B_n| = 2^n \cdot n! \)

The double cosets \( B_n \setminus S_{2n} / B_n \) are indexed by partitions of \( n \)
(Macdonald, Ch. 7)
The induced measure is the *Ewens’s sampling formula*, with $\theta = 1/2$:

$$p_\theta(\lambda) = \frac{n!}{\theta(\theta + 1) \ldots (\theta + n - 1)} \cdot \frac{\theta^{\ell(\lambda)}}{\prod_{i=1}^{n} i^{a_i} a_i!},$$

(2)

where $\ell(\lambda)$ is the number of parts of $\lambda$, and $\lambda$ has $a_i$ parts of size $i$.

**Note:** $\theta = 1$ is the distribution induced by looking at the cycle structure of a uniformly random permutation in $S_n$. 

---

**Ewens measure**
Mapping to partitions

- To each permutation $\sigma \in S_{2n}$ associate a graph $T(\sigma)$ with vertices $1, 2, \ldots, 2n$ and edges $\{\epsilon_i, \epsilon_i^\sigma\}_{i=1}^n$ where $\epsilon_i$ joins vertices $2i-1, 2i$ and $\epsilon_i^\sigma$ joins vertices $\sigma(2i-1), \sigma(2i)$.
- Color the $\epsilon_i$ edges red and the $\epsilon_i^\sigma$ edges as blue.
- Each cycle has an even length. Dividing these cycle lengths by 2 gives a partition of $n$.

Example

Take $n = 3$ and $\sigma = 612543$. The graph $T(\sigma)$ gives $\lambda_\sigma = (2, 1)$. 

Diagram:

- $\epsilon_1$: Connects 1 to 2.
- $\epsilon_2$: Connects 3 to 4.
- $\epsilon_3$: Connects 6 to 5.
- $\epsilon_1^\sigma$: Connects 2 to 6.
- $\epsilon_2^\sigma$: Connects 4 to 3.
- $\epsilon_3^\sigma$: Connects 5 to 4.
Example 3 – Young subgroups of $S_n$

Let $\lambda = (\lambda_1, \lambda_2, \ldots, \lambda_I)$ be a partition of $n$. That is, $
\lambda_1 \geq \lambda_2 \geq \ldots \lambda_I > 0$ and $\sum_i \lambda_i = n$.

The Young subgroup $S_\lambda$ is the set of all permutations in $S_n$ which permute only $\{1, 2, \ldots, \lambda_1\}$ among themselves, only $\{\lambda_1 + 1, \ldots, \lambda_1 + \lambda_2\}$ among themselves, and so on. Thus,

$$S_\lambda \cong S_{\lambda_1} \times S_{\lambda_2} \times \ldots \times S_{\lambda_I}.$$ 

Example

$n = 5, \lambda = (2, 2, 1)$ means only the elements $\{1, 2\}$ can be permuted amongst themselves, $\{3, 4\}$ can be permuted, and $5$ must be fixed:

$$S_\lambda = \{12345, 21345, 12435, 21345\}.$$
$S_\lambda - S_\mu$ double cosets

• Let $\mu = (\mu_1, \ldots, \mu_J)$ be a second partition of $n$.

• The double cosets $S_\lambda \backslash S_n / S_\mu$. can be indexed by contingency tables with fixed row and column sums.

• Contingency tables: $I \times J$ arrays of non-negative integers with row sums $\lambda_1, \ldots, \lambda_I$ and column sums $\mu_1, \ldots, \mu_J$.

This is a standard result, see e.g. James & Kerber, ‘The Representation Theory of the Symmetric Group’, Chapter 1
Contingency tables example

Contingency tables arise from data of a sample of size $n$, classified with two discrete categories. This table has shows 583 subjects classified by 4 levels of eye color and 4 levels of hair color.

<table>
<thead>
<tr>
<th></th>
<th>Black</th>
<th>Brown</th>
<th>Red</th>
<th>Blond</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Brown</td>
<td>68</td>
<td>119</td>
<td>26</td>
<td>7</td>
<td>220</td>
</tr>
<tr>
<td>Blue</td>
<td>20</td>
<td>84</td>
<td>17</td>
<td>94</td>
<td>215</td>
</tr>
<tr>
<td>Hazel</td>
<td>15</td>
<td>54</td>
<td>14</td>
<td>10</td>
<td>93</td>
</tr>
<tr>
<td>Green</td>
<td>5</td>
<td>29</td>
<td>14</td>
<td>16</td>
<td>64</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>108</strong></td>
<td><strong>286</strong></td>
<td><strong>71</strong></td>
<td><strong>127</strong></td>
<td><strong>592</strong></td>
</tr>
</tbody>
</table>

There are $1,225 \times 10^{15}$ tables with these row and column sums.
Mapping from $S_n$ to contingency tables

Let $\lambda = (\lambda_1, \ldots, \lambda_I), \mu = (\mu_1, \ldots, \mu_J)$ be partitions of $n$.

- $\lambda$ defines the sets $L_1 = \{1, \ldots, \lambda_1\}, L_2 = \{\lambda_1 + 1, \ldots, \lambda_1 + \lambda_2\}, \ldots L_I = \{n - \lambda_I + 1, \ldots, n\}$ and analogously $\mu$ defines $M_1, \ldots, M_J$.

- Consider a permutation $\sigma \in S_n$ as an ordered arrangement of the elements $1, 2, \ldots, n$, e.g. $\sigma = 23154$.

Contingency table corresponding to $\sigma$: Let $T_{11}$ be the number of elements from $M_1$ occurring in the positions $L_1$, $T_{12}$ the number of elements from $M_2$ occurring in positions $L_1$, and so on ...

In general, $T_{ij}$ is the number of elements from $M_j$ occurring in positions $L_i$. 
Example

\[ n = 5, \lambda = (3, 2), \mu = (2, 2, 1) \]

\[ L_1 = \{1, 2, 3\}, L_2 = \{4, 5\} \]

\[ M_1 = \{1, 2\}, M_2 = \{3, 4\}, M_3 = \{5\} \]

\[ \sigma = [123]45 \rightarrow \begin{pmatrix} 2 \\ \end{pmatrix} \]

\[ T_{11} - \text{In the first three entries there are 2 elements from } M_1. \]
$n = 5, \lambda = (3, 2), \mu = (2, 2, 1)$

$L_1 = \{1, 2, 3\}, L_2 = \{4, 5\}$

$M_1 = \{1, 2\}, M_2 = \{3, 4\}, M_3 = \{5\}$

$$\sigma = [123]45 \rightarrow \begin{pmatrix} 2 & 1 \end{pmatrix}$$

$T_{12} –$ In the first three entries there is 1 element from $M_2$. 
Example

\[ n = 5, \lambda = (3, 2), \mu = (2, 2, 1) \]

\[ L_1 = \{1, 2, 3\}, L_2 = \{4, 5\} \]

\[ M_1 = \{1, 2\}, M_2 = \{3, 4\}, M_3 = \{5\} \]

\[ \sigma = [123]45 \rightarrow \begin{pmatrix} 2 & 1 & 0 \end{pmatrix} \]

\[ T_{13} \text{ – In the first three entries there, are 0 element from } M_3. \]
Example

\[ n = 5, \lambda = (3, 2), \mu = (2, 2, 1) - \]
\[ L_1 = \{1, 2, 3\}, L_2 = \{4, 5\}, M_1 = \{1, 2\}, M_2 = \{3, 4\}, M_3 = \{5\}. \]
There are five possible tables. Double coset representatives are:

1. \( \sigma = 12345 \rightarrow \begin{pmatrix} 2 & 1 & 0 \\ 0 & 1 & 1 \end{pmatrix} \)
2. \( \sigma = 12534 \rightarrow \begin{pmatrix} 2 & 0 & 1 \\ 0 & 2 & 0 \end{pmatrix} \)
3. \( \sigma = 13425 \rightarrow \begin{pmatrix} 1 & 2 & 0 \\ 1 & 0 & 1 \end{pmatrix} \)
4. \( \sigma = 31542 \rightarrow \begin{pmatrix} 1 & 1 & 1 \\ 1 & 1 & 0 \end{pmatrix} \)
5. \( \sigma = 34512 \rightarrow \begin{pmatrix} 0 & 2 & 1 \\ 2 & 0 & 0 \end{pmatrix} \)
Fisher-Yates distribution

A distribution on $I \times J$ contingency tables with fixed row sums $\lambda_1, \ldots, \lambda_I$ and column sums $\mu_1, \ldots, \mu_J$:

$$P_{\lambda, \mu}(T) = \prod_{j=1}^{J} \binom{T_{1j}, \ldots, T_{ij}}{\mu_j} / \binom{n}{\lambda_1, \ldots, \lambda_I} = \frac{\prod_i (\lambda_i)! \prod_j (\mu_j)!}{n! \prod_{i,j} (T_{ij})!}$$
Fisher-Yates distribution

A distribution on $I \times J$ contingency tables with fixed row sums $\lambda_1, \ldots, \lambda_I$ and column sums $\mu_1, \ldots, \mu_J$:

$$P_{\lambda,\mu}(T) = \prod_{j=1}^{J} \binom{\mu_j}{T_{1j}, \ldots, T_{ij}} / \binom{n}{\lambda_1, \ldots, \lambda_I} = \frac{\prod_i (\lambda_i)! \prod_j (\mu_j)!}{n! \prod_{i,j} (T_{ij})!}$$

Sampling without replacement distribution:

- Suppose that an urn contains $n$ total balls of $J$ different colors, $\mu_j$ of color $j$.
- Make a sequence of $I$ draws without replacement to empty the urn.
- First draw $\lambda_1$ balls, next $\lambda_2$, and so on until there are $\lambda_I = n - \sum_{i=1}^{I-1} \lambda_i$ balls left.
- Set $T_{ij}$ to be the number of color $j$ in the $i$th draw.
Induced by uniform distribution on $S_n$

The sampling without replacement scheme generates a contingency table with the Fisher-Yates distribution.

It also proves that uniform on $S_n$ induces the Fisher-Yates: Color elements according to $\mu$, then $\sigma$ records the overall sequence of draws.

$\lambda = (3, 2), \mu = (2, 2, 1)$: $M_1 = \{1, 2\}, M_2 = \{3, 4\}, M_3 = \{5\}$.

$\sigma = [134] [25] \rightarrow \begin{pmatrix} 1 & 2 & 0 \\ 1 & 0 & 1 \end{pmatrix}$

Note: When $I = J = 2$, this is the hypergeometric distribution. When $I = 2$, $J > 2$, it is the multivariate hypergeometric distribution.
Induced by uniform distribution on $S_n$

The sampling without replacement scheme generates a contingency table with the Fisher-Yates distribution.

It also proves that uniform on $S_n$ induces the Fisher-Yates: Color elements according to $\mu$, then $\sigma$ records the overall sequence of draws.

$\lambda = (3, 2), \mu = (2, 2, 1): M_1 = \{1, 2\}, M_2 = \{3, 4\}, M_3 = \{5\}$.

$$\sigma = [1\, 3\, 4][2\, 5] \rightarrow \begin{pmatrix} 1 & 2 & 0 \\ 1 & 0 & 1 \end{pmatrix}$$

**Note:** When $I = J = 2$, this is the hypergeometric distribution. When $I = 2, J > 2$, it is the multivariate hypergeometric distribution.
Independence model conditioned on row and column sums

• Create a table by dropping $n$ balls in $I \cdot J$ cells.

• Assume that column traits and row traits are independent: A ball is dropped in cell $i, j$ with probability $p_i \cdot q_j$.

• Under this model,

$$
P(T) = \frac{n!}{\prod_{i,j} T_{ij}!} \cdot \prod_{i,j} (p_i \cdot q_j)^{T_{ij}} = \frac{n!}{\prod_{i,j} T_{ij}!} \prod_{i=1}^I p_i^{\lambda_i} \prod_{j=1}^J q_j^{\mu_j},$$

if $T$ has row sums $\lambda_i$ and column sums $\mu_j$.

Thus, a table generated under the independence model **conditioned on the row and column sums** has the Fisher-Yates distribution (regardless of $p_i, q_j$).
The Fisher-Yates distribution is used to test the hypotheses that two traits (e.g. hair color and eye color) are independent (The chi-squared statistic has limiting chi-squared distribution.)

The table with the highest probability (the largest double coset) is the one closest to the ‘independence table’:

\[ T_{ij}^* = \frac{\lambda_i \cdot \mu_j}{n}. \]

Dyer, Kannan, Mount, 1994 + Diaconis & Gangolli, 1994 – With \( \lambda, \mu \) as parameters, counting the number of contingency tables is \# -P complete.
A Markov chain on contingency tables

Outline:

1. Random transpositions on $S_n$
2. Induced random transpositions chain on tables
3. Mixing time overview + Motivation
4. Eigenvalues/eigenfunctions of Markov chains
5. Results

Markov chain and mixing time definitions and standard results from ‘Markov Chains and Mixing Times”, Levin, Peres, & Wilmer
Random transpositions on $S_n$

- Think of a permutation as a deck of cards: Pick one card with your left hand, one with your right hand, and swap the cards (allowing for the possibility of picking the same card).

- Transition matrix for the random transpositions Markov chain:

$$P(x, y) = \begin{cases} 
\frac{2}{n^2} & \text{if } y = (ij)x, \ i \neq j \\
\frac{1}{n} & \text{if } y = x \\
0 & \text{else}
\end{cases}.$$
**Contingency table chain**

Let $\lambda, \mu$ be partitions of $n$ and $T_{\lambda, \mu}$ be the space of contingency tables with row/column sums given by $\lambda$ and $\mu$. For a permutation $x \in S_n$, let $T^x$ denote the contingency table (double coset) induced by $x$.

Random transpositions Markov chain on the space of tables $T_{\lambda, \mu}$:

1. From a table $T$, pick an $x \in S_n$ such that $T^x = T$.

2. Choose $y$ from $P(x, \cdot)$, then move to $T' = T^y$. 
Contingency table chain

Let $\lambda, \mu$ be partitions of $n$ and $\mathcal{T}_{\lambda,\mu}$ be the space of contingency tables with row/column sums given by $\lambda$ and $\mu$. For a permutation $x \in S_n$, let $T^x$ denote the contingency table (double coset) induced by $x$.

Random transpositions Markov chain on the space of tables $\mathcal{T}_{\lambda,\mu}$:

1. From a table $T$, pick an $x \in S_n$ such that $T^x = T$.
2. Choose $y$ from $P(x, \cdot)$, then move to $T' = T^y$.

The transition probability doesn’t depend on the choice of double coset representative $x$ (Dynkin’s condition), so this is a well-defined Markov chain with Fisher-Yates as stationary distribution.
Dynkin’s condition

“Lumped” Markov chain:

Lemma

Suppose $X_t$ is a Markov chain on a state space $\Omega$ with transition matrix $P$ and stationary distribution $\pi$, and $\sim$ is an equivalence relation on $\Omega$ with equivalence classes $\tilde{\Omega} = \{[x] : x \in \Omega\}$. Assume $P$ satisfies

$$P(x, [y]) := \sum_{y' \sim y} P(x, y) = P(x', [y])$$

whenever $x \sim x'$. Then $[X_t]$ is a Markov chain on the space $\tilde{\Omega}$ and transition matrix $\tilde{P}([x], [y]) := P(x, [y])$.

The stationary distribution of the lumped chain is $\tilde{\pi}([x]) = \sum_{x' \sim x} \pi(x)$. 
Example

\(\lambda = (3, 2), \mu = (2, 1, 1):\)

\[ T^x = \begin{pmatrix} 2 & 1 & 0 \\ 0 & 1 & 1 \end{pmatrix} \quad \rightarrow \quad T^y = \begin{pmatrix} 1 & 2 & 0 \\ 1 & 0 & 1 \end{pmatrix} \]

\(x = 12345\) \quad \rightarrow \quad y = 14325
\( \lambda = (3, 2), \mu = (2, 1, 1): \)

\[
T^x = \begin{pmatrix} 2 & 1 & 0 \\ 0 & 1 & 1 \end{pmatrix} \quad \rightarrow \quad T^y = \begin{pmatrix} 1 & 2 & 0 \\ 1 & 0 & 1 \end{pmatrix}
\]

\( x = 12345 \quad \rightarrow \quad y = 14325 \)

Note that

\[
T^y = T^x + \begin{pmatrix} -1 & 1 & 0 \\ 1 & -1 & 0 \end{pmatrix}
\]
Transition probabilities

What are the possible moves in the table?

- Let $F_{(i_1,j_1),(i_2,j_2)}(T) = T'$ denote the table with

$$
T'_{i_1,j_1} = T_{i_1,j_1} - 1 \\
T'_{i_2,j_2} = T_{i_2,j_2} - 1 \\
T'_{i_1,j_2} = T_{i_1,j_2} + 1 \\
T'_{i_2,j_1} = T_{i_2,j_1} + 1,
$$

and all other entries of $T'$ the same as $T$.

- For $i_1 \neq i_2, j_1 \neq j_2$:

$$
\tilde{P}(T, F_{(i_1,j_1),(i_2,j_2)}(T)) = \frac{T_{i_1,j_1} \cdot T_{i_2,j_2}}{n^2}.
$$
Example again

\( \lambda = (3, 2), \mu = (2, 1, 1) : \)

\[
T^x = \begin{pmatrix} 2 & 1 & 0 \\ 0 & 1 & 1 \end{pmatrix} \quad \rightarrow \quad T^y = \begin{pmatrix} 1 & 2 & 0 \\ 1 & 0 & 1 \end{pmatrix}
\]

\( x = 12345 \quad \rightarrow \quad y = 14325 \quad y = 42315 \)

This transition has probability:

\[
\frac{T^x_{11} \cdot T^x_{22}}{5^2} = \frac{2}{25}.
\]
Consider another Markov chain with the same type of moves, but different probabilities:

- Pick two table entries \((i_1, j_1), (i_2, j_2)\) uniformly. If allowed, make the same swap move as before.
- Symmetric chain: If \(i_1 \neq i_2, j_1 \neq j_2\) and \(T_{i_1,j_1}, T_{i_2,j_2} > 0\), then

\[
\hat{P}(T, F_{(i_1,j_1),(i_2,j_2)}(T)) = \frac{1}{I \cdot J}.
\]
Uniform distribution chain

Consider another Markov chain with the same type of moves, but different probabilities:

- Pick two table entries \((i_1, j_1), (i_2, j_2)\) uniformly. If allowed, make the same swap move as before.
- Symmetric chain: If \(i_1 \neq i_2, j_1 \neq j_2\) and \(T_{i_1,j_1}, T_{i_2,j_2} > 0\), then
  \[
  \hat{P}(T, F_{(i_1,j_1),(i_2,j_2)}(T)) = \frac{1}{I \cdot J}.
  \]

This Markov chain has \textit{uniform stationary distribution}.

- It has been studied as a way of sampling from the uniform distribution and approximately counting the number of tables – Introduced by Diaconis and Gangolli, 1995
Mixing times

How many steps do we need to run the Markov chain before it reaches its stationary distribution?
How many steps do we need to run the Markov chain before it reaches its stationary distribution?

Suppose \((X_t)_{t \geq 0}\) is a discrete-time Markov chain on a finite state space \(\Omega\) with transition matrix \(P\) and stationary distribution \(\pi\).

- **Total variation distance:**

  \[
d(t) := \sup_{x \in \Omega} \| P^t(x, \cdot) - \pi(\cdot) \|_{TV}
  = \sup_{x \in \Omega} \max_{A \subset \Omega} | P(X_t \in A \mid X_0 = x) - \pi(A) |.
  \]

- **The mixing time** is

  \[
t_{mix} := \inf \{ t > 0 : d(t) < 1/4 \}.
  \]
Eigenvalues of Markov chains

- Let $1 = \beta_0 > \beta_1 \geq \ldots \geq \beta_{|\Omega|-1} \geq -1$ be the eigenvalues of $P$ with corresponding eigenfunctions $f_0, f_1, \ldots, f_{|\Omega|-1}$. That is, $f_i : \Omega \rightarrow \mathbb{R}$ and

$$E[f_i(X_1) \mid X_0 = x] = \beta_i \cdot f_i(x), \quad x \in \Omega.$$
Eigenvalues of Markov chains

• Let $1 = \beta_0 > \beta_1 \geq \ldots \geq \beta_{|\Omega|-1} \geq -1$ be the eigenvalues of $P$ with corresponding eigenfunctions $f_0, f_1, \ldots, f_{|\Omega|-1}$. That is, $f_i : \Omega \to \mathbb{R}$ and

$$E[f_i(X_1) \mid X_0 = x] = \beta_i \cdot f_i(x), \quad x \in \Omega.$$ 

• We can choose the eigenfunctions $\{f_i\}_{i \geq 0}$ to be *orthonormal* with respect to the stationary distribution:

$$\sum_{x \in \Omega} f_i(x) \cdot f_j(x) \cdot \pi(x) = 1(i = j).$$

• This is an *orthonormal basis* of the inner product space $\ell^2(\pi) = (\mathbb{R}^\Omega, \langle \cdot, \cdot \rangle_\pi)$, where

$$\langle f, g \rangle_\pi := \sum_{x \in \Omega} f(x)g(x)\pi(x).$$
The eigenvalues and eigenfunctions tell us a lot about $P$.

- **Chi-square distance:**

$$\| P^t(x, \cdot) - \pi(\cdot) \|_2^2 := \sum_{y \in \Omega} \frac{|P^t(x, y) - \pi(y)|^2}{\pi(y)} \quad = \sum_{i=1}^{\Omega-1} \beta_i^{2t} \cdot f_i^2(x).$$

- **Total variation bound:**

$$\| P^t(x, \cdot) - \pi(\cdot) \|_{TV}^2 \leq \frac{1}{4} \| P^t(x, \cdot) - \pi(\cdot) \|_2^2 = \frac{1}{4} \sum_{i=1}^{\Omega-1} \beta_i^{2t} \cdot f_i^2(x).$$
Eigenvalues of lumped chain

Suppose $P$ is a Markov chain on $\Omega$ and $\tilde{P}$ is the lumped chain on equivalence classes $\tilde{\Omega} = \{[x] : x \in \Omega\}$.

1. If $f : \Omega \rightarrow \mathbb{R}$ is an eigenfunction of $P$ with eigenvalue $\lambda$ which is constant on equivalence classes then the natural projection $\tilde{f} : \tilde{\Omega} \rightarrow \mathbb{R}$ (defined $\tilde{f}([x]) = f(x)$) is an eigenfunction of $\tilde{P}$ with eigenvalue $\lambda$.

2. Conversely, if $g : \tilde{\Omega} \rightarrow \mathbb{R}$ is an eigenfunction of $\tilde{P}$ with eigenvalue $\lambda$, then its lift $\check{g} : \Omega \rightarrow \mathbb{R}$ (defined $\check{g}(x) = g([x])$) is an eigenfunction of $P$ with eigenvalue $\lambda$.
Eigenvalues of random transpositions

Diaconis & Shashahani, 1981: If $P$ is the random transpositions chain on $S_n$, then

- $P$ has an eigenvalue $\beta_\rho$ for each partition $\rho = (\rho_1, \rho_2, \ldots, \rho_k)$ of $n$.

- Formula:

  $$\beta_\rho = \frac{1}{n} + \frac{1}{n^2} \sum_{j=1}^{k} [(\rho_j - j)(\rho_j - j + 1) - j(j - 1)]$$

- The multiplicity of $\beta_\rho$ is $d_\rho^2$, where $d_\rho$ is the hook-length of the partition $\rho$. 

So we know eigenvalues of the contingency table are some subset of these. Which partitions show up, and with what multiplicity? What are the eigenfunctions?
Eigenvalues of random transpositions

Diaconis & Shashahani, 1981: If $P$ is the random transpositions chain on $S_n$, then

- $P$ has an eigenvalue $\beta_\rho$ for each partition $\rho = (\rho_1, \rho_2, \ldots, \rho_k)$ of $n$.

- Formula:

$$\beta_\rho = \frac{1}{n} + \frac{1}{n^2} \sum_{j=1}^{k} [(\rho_j - j)(\rho_j - j + 1) - j(j - 1)]$$

- The multiplicity of $\beta_\rho$ is $d_\rho^2$, where $d_\rho$ is the hook-length of the partition $\rho$.

So we know eigenvalues of the contingency table are some subset of these. Which partitions show up, and with what multiplicity? What are the eigenfunctions?
**Theorem**

Let $\lambda = (\lambda_1, \ldots, \lambda_I)$, $\mu = (\mu_1, \ldots, \mu_J)$ be partitions of $n$ and $\tilde{P}$ the random transpositions Markov chain on the space of contingency tables $T_{\lambda,\mu}$. Suppose $\beta$ is an eigenvalue. Then for some $0 \leq m \leq \lfloor n/2 \rfloor$,

$$\beta = 1 - \frac{2m(n + 1 - m)}{n^2}.$$ 

The eigenbasis for $\beta$ is the set of orthogonal polynomials of the Fisher-Yates distribution of degree $m$. 

**Remarks:**

1. This corresponds to the two-part partition $(n - m, m)$. Thus, there are a maximum of $\lfloor n/2 \rfloor + 1$ unique eigenvalues.

2. We do not know (yet) the multiplicity of each eigenvalue. This would come from the number of orthogonal polynomials of degree $m$. 
Theorem
Let $\lambda = (\lambda_1, \ldots, \lambda_l), \mu = (\mu_1, \ldots, \mu_J)$ be partitions of $n$ and $\tilde{P}$ the random transpositions Markov chain on the space of contingency tables $\mathcal{T}_{\lambda,\mu}$. Suppose $\beta$ is an eigenvalue. Then for some $0 \leq m \leq \lfloor n/2 \rfloor$,

$$\beta = 1 - \frac{2m(n + 1 - m)}{n^2}.$$ 

The eigenbasis for $\beta$ is the set of orthogonal polynomials of the Fisher-Yates distribution of degree $m$.

Remarks:
1. This corresponds to the two-part partition $(n - m, m)$. Thus, there are a maximum of $\lfloor n/2 \rfloor + 1$ unique eigenvalues.
2. We do not know (yet) the multiplicity of each eigenvalue. This would come from the number of orthogonal polynomials of degree $m$. 
Orthogonal polynomials

Let $\pi$ be a probability distribution on a finite space $\Omega \subset \mathbb{N}^d$. Let $\ell^2(\pi)$ be the space of functions $f : \Omega \to \mathbb{R}$ with the inner product

$$\langle f, g \rangle_\pi = \mathbb{E}_\pi[f(X)g(X)] = \sum_{x \in \Omega} f(x)g(x)\pi(x).$$

**Definition:** A set of functions $\{q_m\}_{0 \leq m < |\Omega|}$ are orthogonal in $\ell^2(\pi)$ if

$$\langle q_m, q_\ell \rangle_\pi = d_m^2 \mathbf{1}(m = \ell).$$

**Fact:** There exists a basis of $\ell^2(\pi)$ of orthogonal polynomials.

For the multivariate hypergeometric distribution, these are the Hahn polynomials.
Second-largest eigenvalue

Corollary

The second-largest eigenvalue is

\[ \beta_1 = 1 - \frac{2}{n} \]

and has multiplicity \((I - 1) \cdot (J - 1)\). A basis for the space of eigenfunctions of \(\beta_1\) is given by

\[ f_{ij}(x) = x_{ij} - \frac{\lambda_i \cdot \mu_j}{n}, \quad 1 \leq i < I, 1 \leq j < J. \]

Note the formula for the eigenvalue corresponds to the partition \((n - 1, 1)\).
For $2 \times J$ tables, the stationary distribution is multi-variate hypergeometric, and we know more:

**Theorem**

Let $\lambda = (n - k, k)$ for some $k \leq \lfloor n/2 \rfloor$, $\mu = (\mu_1, \ldots, \mu_J)$, and $\tilde{P}$ be the random swap Markov chain on the space of contingency tables $\mathcal{T}_{\lambda, \mu}$. Then

$$\beta_m = 1 - \frac{2m(n + 1 - m)}{n^2}, \quad 0 \leq m \leq k,$$

in an eigenvalue with eigenbasis the orthogonal polynomials of degree $m$ for the multivariate hypergeometric distribution. The multiplicity of $\beta_m$ is the size of the set

$$\left\{ (x_1, \ldots, x_{J-1}) \in \mathbb{N}^{J-1} : \sum x_j = m, x_j < \mu_{J-j+1} \right\}.$$
Criterion for orthogonal polynomials as eigenfunctions

Khare & Zhou, 2009: Suppose $\pi$ is a univariate distribution on $\Omega$ and $\{q_m\}_{0 \leq m < |\Omega|}$ is an orthogonal basis of $\ell^2(\pi)$.

• Let $(X_t)_{t \geq 0}$ be a reversible Markov chain with transition matrix $P$ and stationary distribution $\pi$.

• Suppose for $0 \leq m < |\Omega|$,

\[ E[X_1^m \mid X_0 = x] = \beta_m x^m + (\text{terms in } x \text{ of degree } < m). \]

• Then $P$ has eigenvalue $\beta_m$ with eigenfunction $q_m$.

Same result holds for multivariate distributions, except eigenvalue $\beta_m$ will have eigenbasis of orthogonal functions with degree equal to $m$. 
Let \((X_t)_{t \geq 0}\) be the Markov chain, \(X_t^{ij}\) the entry in the \((i, j)\) cell. Then,

\[
\mathbb{P}
\left( X_{t+1}^{ij} = X_t^{ij} + 1 \mid X_t = x \right)
= \frac{2}{n^2} \cdot \sum_{k \neq i, \ell \neq j} x_{kj} x_{i\ell}
= \frac{2}{n^2} \cdot (\lambda_i - x_{ij})(\mu_j - x_{ij})
\]

\[
\mathbb{P}
\left( X_{t+1}^{ij} = X_t^{ij} - 1 \mid X_t = x \right)
= \frac{2}{n^2} \cdot \sum_{k \neq i, \ell \neq j} x_{ij} x_{k\ell}
= \frac{2}{n^2} \cdot x_{ij}(n - \lambda_i - \mu_j + x_{ij})
\]
Let \((X_t)_{t \geq 0}\) be the Markov chain, \(X_t^{ij}\) the entry in the \((i, j)\) cell. Then,

\[
\begin{align*}
\mathbb{P}\left( X_{t+1}^{ij} = X_t^{ij} + 1 \mid X_t = x \right) &= \frac{2}{n^2} \cdot \sum_{k \neq i, \ell \neq j} x_{kj} x_{i\ell} \\
&= \frac{2}{n^2} \cdot (\lambda_i - x_{ij})(\mu_j - x_{ij})
\end{align*}
\]

\[
\begin{align*}
\mathbb{P}\left( X_{t+1}^{ij} = X_t^{ij} - 1 \mid X_t = x \right) &= \frac{2}{n^2} \cdot \sum_{k \neq i, \ell \neq j} x_{ij} x_{k\ell} \\
&= \frac{2}{n^2} \cdot x_{ij}(n - \lambda_i - \mu_j + x_{ij})
\end{align*}
\]

Using this, we can easily calculate \(\mathbf{E}\left[(X_{t+1}^{ij})^m \mid X_t = x\right].\)
Total variation bound?

Recall: If \( \{f_i\}_{i \geq 1} \) are orthonormal eigenfunctions, then

\[
\| P^t(x, \cdot) - \pi(\cdot) \|_{TV}^2 \leq \frac{1}{4} \| P^t(x, \cdot) - \pi(\cdot) \|_2^2 = \frac{1}{4} \sum_{i=1}^{\Omega - 1} \beta_i^{2t} \cdot f_i^2(x) \quad (3)
\]

- We know the eigenfunctions are orthogonal polynomials.

- For 2 \( \times \) J tables, these are the orthogonal polynomials for the multi-variate hypergeometric distribution. These are explicitly known (e.g. Griffiths, 2006). For special cases for \( x \), Eqn. (3) can be simplified.

- **Problem:** For mixing time, we need to be able to analyze Equation (3) for every \( x \in \Omega \).
Upper bound result for $2 \times J$ tables

- $\lambda = (n - k, k), \mu = (\mu_1, \ldots, \mu_J)$, for $k \leq \lfloor n/2 \rfloor$.
  
  **Assumption:** $\mu_j > k$

- For $j = 1, \ldots, J$, let $k\mathbf{e}_j$ be the table with the second row all 0 except $k$ in the $j$th column.

- **Example:** $n = 10, \mu = (4, 3, 3), \lambda = (8, 2)$

\[
2\mathbf{e}_1 = \begin{pmatrix} 2 & 3 & 3 \\ 2 & 0 & 0 \end{pmatrix}
\]
Upper bound result for $2 \times J$ tables

- $\lambda = (n - k, k), \mu = (\mu_1, \ldots, \mu_J)$, for $k \leq \lfloor n/2 \rfloor$.
  
  **Assumption:** $\mu_j > k$

- For $j = 1, \ldots, J$, let $ke_j$ be the table with the second row all 0 except $k$ in the $j$th column.

- **Example:** $n = 10, \mu = (4, 3, 3), \lambda = (8, 2)$

  \[
  2e_1 = \begin{pmatrix} 2 & 3 & 3 \\ 2 & 0 & 0 \end{pmatrix}
  \]

**Lemma**

*For any $c > 0$, we get* $\| \tilde{P}^t(ke_j, \cdot) - \pi(\cdot) \|_2^2 \leq e^{-c}$ *for*

\[
t \geq \frac{n}{4} \left( \log \left( \frac{k \cdot n \cdot (n - \mu_j)}{(n - k) \cdot (\mu_j - k)} \right) + c \right).
\]
Special cases:

1. \( I = J = 2 \) and \( k = n/2 - 1, \mu = n/2 \), then
   \[
   \frac{k \cdot n \cdot (n - \mu_i)}{(n - k) \cdot (\mu_j - k)} \sim n
   \]

2. \( k \) is fixed and small, \( \mu_j = n/J \), then
   \[
   \frac{k \cdot n \cdot (n - \mu_i)}{(n - k) \cdot (\mu_j - k)} \sim k \cdot (J - 1)
   \]

Remark: Random transpositions on \( S_n \) mixes in time \( (n/2) \log(n) \).
Mixing time lower bound

Recall the eigenvalue $\beta = 1 - 2/n$ has explicit eigenfunctions:

$$f_{ij}(x) = x_{ij} - \frac{\lambda_i \cdot \mu_j}{n}$$

Using these, we can apply Wilson’s method for a lower bound.

**Lemma**

Let $\lambda = (\lambda_1, \ldots, \lambda_I), \mu = (\mu_1, \ldots, \mu_J)$ be any partitions of $n$. For any $i, j$ such that $n > 2(\lambda_i + \mu_j)$, there is a $c > 0$ such that

$$t_{mix} \geq \frac{n}{4} \left( \log \left( \min(\lambda_i, \mu_j) - \frac{\lambda_i \mu_j}{n} \right) + c \right).$$
• Double cosets of groups can be enumerated by interesting combinatorial objects and the uniform distribution on the group can induce nice distributions on double cosets.

• Markov chains on the original group can induce Markov chains on the space of double cosets.

• Random transposition Markov chain on contingency tables: Orthogonal polynomial eigenfunctions for the Fisher-Yates distribution.
Thank you!

Questions?