Markov Chains from Jeu de Taquin

Arvind Ayyer
(joint work with Steve Klee and Anne Schilling)

International Conference on Limit Theorems and Probability

Indian Institute of Science Bangalore

January 9, 2013

Outline

1. Tsetlin Library
2. Jeu de Taquin
3. Promotion on posets
4. Markov chain
5. Stationary distribution
6. Eigenvalues
7. Proof ideas

A model of a library

- n books on a shelf

$$\boxed{B_1}\boxed{B_2}\cdots\boxed{B_n}$$

A model of a library

- n books on a shelf

$$\boxed{B_1}\boxed{B_2}\cdots\boxed{B_n}$$

- The probability of choosing book B_i is x_i.
- Once the book is chosen, it is moved to the back.

$$\boxed{B_1}\boxed{B_2}\cdots\boxed{B_i}\cdots\boxed{B_n} \to \boxed{B_1}\boxed{B_2}\cdots\boxed{B_n}\boxed{B_i} \text{ with probability } x_i.$$

A Markov chain on permutations

- Let $\pi \in S_n$ be a permutation.
- The stationary distribution is given by [Hendricks '72]

$$\mathbb{P}(\pi) = \prod_{i=1}^{n} \frac{x_{\pi_i}}{x_{\pi_1} + \cdots + x_{\pi_i}}.$$

A Markov chain on permutations

- Let $\pi \in S_n$ be a permutation.
- The stationary distribution is given by [Hendricks '72]

$$\mathbb{P}(\pi) = \prod_{i=1}^{n} \frac{x_{\pi_i}}{x_{\pi_1} + \cdots + x_{\pi_i}}.$$

- A *derangement* is a permutation with no fixed points.
- d_m be the number of derangements in S_m.
- Let M_n be the transition matrix or generator. Then [Phatarfod '91]

$$\det(M_n - \lambda \mathbb{1}) = \prod_{S \subset [n]} (\lambda + x_S)^{d_{|S|}}$$

where $x_S = \sum_{i \in S} x_i$.

Example

- The case of $n = 3$:

$$M_3 = \begin{pmatrix} * & x_3 & 0 & 0 & x_3 & 0 \\ x_2 & * & x_2 & 0 & 0 & 0 \\ 0 & 0 & * & x_3 & 0 & x_3 \\ x_1 & 0 & x_1 & * & 0 & 0 \\ 0 & 0 & 0 & x_2 & * & x_2 \\ 0 & x_1 & 0 & 0 & x_1 & * \end{pmatrix} \begin{bmatrix} 123 \\ 132 \\ 213 \\ 231 \\ 312 \\ 321 \end{bmatrix}$$

- $$\mathbb{P}(231) = \frac{x_3 x_1}{(x_2 + x_3)(x_1 + x_2 + x_3)}$$

- Eigenvalues: 0, $-x_1 - x_2$, $-x_1 - x_3$, $-x_2 - x_3$ and $-x_1 - x_2 - x_3$ twice.

Generalizations

- Umpteen generalizations!
- Different moves, more shelves.
- Infinite libraries.
- Hyperplane arrangements [Bidigare, Hanlon, Rockmore '99]
- Left regular bands (monoids) [Brown '00]

Standard Young Tableaux

- A Young diagram is a representation of a partition $\lambda = (\lambda_1, \lambda_2, \ldots, \lambda_k)$ where the entries are weakly decreasing.
- Originally defined by Schützenberger on skew tableaux
- Bijection on SYT

1	3	4	7
2	5		
6	8		

Standard Young Tableaux

- A Young diagram is a representation of a partition $\lambda = (\lambda_1, \lambda_2, \ldots, \lambda_k)$ where the entries are weakly decreasing.
- Originally defined by Schützenberger on skew tableaux
- Bijection on SYT

·	3	4	7
2	5		
6	8		

Standard Young Tableaux

- A Young diagram is a representation of a partition $\lambda = (\lambda_1, \lambda_2, \ldots, \lambda_k)$ where the entries are weakly decreasing.
- Originally defined by Schützenberger on skew tableaux
- Bijection on SYT

2	3	4	7
·	5		
6	8		

Standard Young Tableaux

- A Young diagram is a representation of a partition $\lambda = (\lambda_1, \lambda_2, \ldots, \lambda_k)$ where the entries are weakly decreasing.
- Originally defined by Schützenberger on skew tableaux
- Bijection on SYT

2	3	4	7
5	·		
6	8		

Standard Young Tableaux

- A Young diagram is a representation of a partition $\lambda = (\lambda_1, \lambda_2, \ldots, \lambda_k)$ where the entries are weakly decreasing.
- Originally defined by Schützenberger on skew tableaux
- Bijection on SYT

2	3	4	7
5	8		
6	·		

Standard Young Tableaux

- A Young diagram is a representation of a partition $\lambda = (\lambda_1, \lambda_2, \ldots, \lambda_k)$ where the entries are weakly decreasing.
- Originally defined by Schützenberger on skew tableaux
- Bijection on SYT

2	3	4	7
5	8		
6	9		

Standard Young Tableaux

- A Young diagram is a representation of a partition $\lambda = (\lambda_1, \lambda_2, \ldots, \lambda_k)$ where the entries are weakly decreasing.
- Originally defined by Schützenberger on skew tableaux
- Bijection on SYT

1	2	3	6
4	7		
5	8		

Standard Young Tableaux

- A Young diagram is a representation of a partition $\lambda = (\lambda_1, \lambda_2, \ldots, \lambda_k)$ where the entries are weakly decreasing.
- Originally defined by Schützenberger on skew tableaux
- Bijection on SYT

$$\begin{array}{|c|c|c|c|}\hline 1 & 2 & 3 & 6 \\ \hline 4 & 7 \\ \cline{1-2} 5 & 8 \\ \cline{1-2}\end{array} = \partial_1 \left(\begin{array}{|c|c|c|c|}\hline 1 & 3 & 4 & 7 \\ \hline 2 & 5 \\ \cline{1-2} 6 & 8 \\ \cline{1-2}\end{array} \right)$$

A Minor modification - ∂_j

- First used by Edelman, Hibi and Stanley.
- Instead of starting by removing 1, we can remove integer j.
- Continue the procedure the same way.
- Add $n + 1$ at the end
- Subtract 1 from everything larger than j.

Example

- $$\partial_4 \left(\begin{array}{|c|c|c|c|} \hline 1 & 2 & 3 & 6 \\ \hline 4 & 7 \\ \cline{1-2} 5 & 8 \\ \cline{1-2} \end{array} \right)$$

Example

1	2	3	6
·	7		
5	8		

Example

1	2	3	6
5	7		
·	8		

Example

-

1	2	3	6
5	7		
8	·		

Example

1	2	3	6
5	7		
8	9		

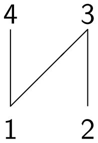
Example

1	2	3	5
4	6		
7	8		

Posets

- P - a partially ordered set with order $\prec$
- $|P| = n$, "naturally" labeled by integers in $[n]$.
- $\mathcal{L}(P)$ - linear extensions of P, ways of arranging elements of P in a line respecting the order

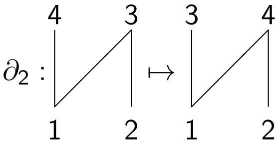
$$\mathcal{L}(P) = \{\pi \in S_n : i \prec j \Rightarrow \pi_i^{-1} < \pi_j^{-1}\} \ni e$$

- Eg: $P =$ 

, $\mathcal{L}(P) = \{1234, 1243, 1423, 2134, 2143\}$.

Jeu de Taquin (aka Promotion)

- The action of ∂_i is exactly as before.
- For example,



- Can be used to define a Markov Chain

Relation with Transposition

- Define τ_i on $\mathcal{L}(P)$

$$\pi\tau_i = \begin{cases} \pi_1 \cdots \pi_{i-1}\pi_{i+1}\pi_i \cdots \pi_n & \text{if } \pi_i \text{ and } \pi_{i+1} \text{ are not comparable in } P, \\ \pi_1 \cdots \pi_n & \text{otherwise.} \end{cases}$$

- Then [Haiman '92, Malvenuto & Reutenauer '94]

$$\partial_j = \tau_n \tau_{n-1} \cdots \tau_j.$$

The Directed Graph

- Given P, let G be the graph whose vertex set is $\mathcal{L}(P)$
- There is an edge $\pi \to \pi'$ if $\pi' = \partial_j(\pi)$ for some j.

Lemma

G is strongly connected.

- We will define two Markov chains on this underlying graph

Uniform Promotion Graph

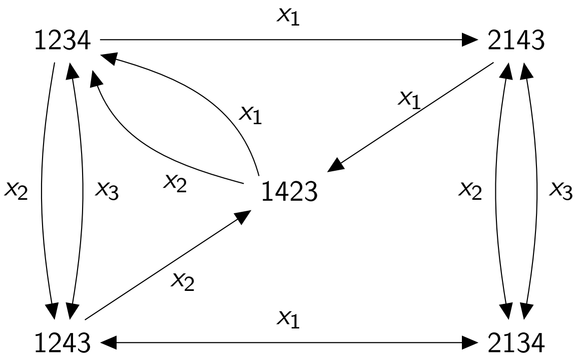
- The edge $\pi \to \pi'$, where $\pi' = \partial_j(\pi)$ has weight x_j.
- In this Markov chain, we give a probability distribution to the ∂_i's.

Theorem

The stationary distribution of the Markov chain is uniform.

- Follows from the fact that $\partial_i^k = \partial_i$ for large enough k.

Example



Promotion Graph

- The edge $\pi \to \pi'$, where $\pi' = \partial_j(\pi)$ has weight $x_{\pi(j)}$.
- In this Markov chain, we give a probability distribution to the values of the current state π.
- The stationary distribution of this Markov chain is no longer uniform.

Stationary Distribution

Theorem (1)

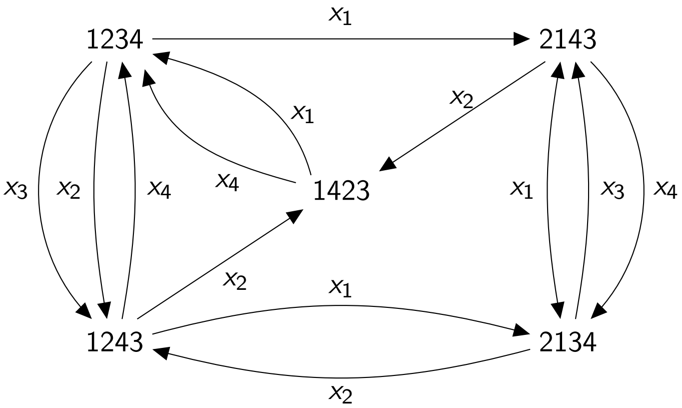
The stationary state weight $w(\pi)$ of the linear extension $\pi \in \mathcal{L}(P)$ for the continuous time Markov chain for the promotion graph is given by

$$w(\pi) = \prod_{i=1}^{n} \frac{x_1 + \cdots + x_i}{x_{\pi_1} + \cdots + x_{\pi_i}} ,$$

assuming $w(e) = 1$.

Note that $w(\pi)$ is independent of P.
Proved by induction.

Example



Example (contd.)

The transition matrix this time is given by

$$\begin{pmatrix} * & x_4 & x_1 + x_4 & 0 & 0 \\ x_2 + x_3 & * & 0 & x_2 & 0 \\ 0 & x_2 & * & 0 & x_2 \\ 0 & x_1 & 0 & * & x_1 + x_4 \\ x_1 & 0 & 0 & x_1 + x_3 & * \end{pmatrix}$$

Notice that row sums are no longer zero. The stationary distribution is

$$\left(1, \quad \frac{x_1 + x_2 + x_3}{x_1 + x_2 + x_4}, \quad \frac{(x_1 + x_2)(x_1 + x_2 + x_3)}{(x_1 + x_2)(x_1 + x_2 + x_4)}, \quad \frac{x_1}{x_2}, \quad \frac{x_1(x_1 + x_2 + x_3)}{x_2(x_1 + x_2 + x_4)}\right).$$

Recall Tsetlin library!

Special Posets

- A **rooted tree** is a connected poset, where each node has at most one successor.
- A **rooted forest** is a union of rooted trees.
- A **chain** is a totally ordered set.
- A **union of chains** is also a rooted forest.

Rooted Forests

- Note that $\sum_\pi w(\pi) \neq 1$ in general.
- The **partition function** Z_P is the prefactor that makes $\mathbb{P}(\pi) = w(\pi)/Z_P$ a probability distribution.

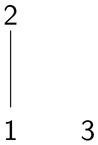
Theorem (2)

Let P be a rooted forest of size n and let $x_{\preceq i} = \sum_{j \preceq i} x_j$. The partition function for the promotion graph is given by

$$Z_P = \prod_{i=1}^{n} \frac{x_1 + \cdots + x_i}{x_{\preceq i}}.$$

Example

- $\mathcal{L}(P) = \{123, 132, 312\}$

$P =$ 

- $$w(123) = 1,\ w(132) = \frac{(x_1 + x_2)}{(x_1 + x_3)},\ w(312) = \frac{x_1(x_1 + x_2)}{x_3(x_1 + x_3)}$$

- $Z_P = \frac{x_1(x_1+x_2)(x_1+x_2+x_3)}{x_1(x_1+x_2)x_3}$

$$\mathbb{P}(123) = \frac{x_3}{x_1 + x_2 + x_3},\ \mathbb{P}(132) = \frac{x_3(x_1 + x_2)}{(x_1 + x_3)(x_1 + x_2 + x_3)},$$

$$\mathbb{P}(312) = \frac{x_1(x_1 + x_2)}{(x_1 + x_3)(x_1 + x_2 + x_3)}$$

More terminology

- An **upper set** S in P is a subset of $[n]$ such that if $x \in S$ and $y \succeq x$, then also $y \in S$.
- Let L be the lattice (by inclusion) of upper sets in P.
- $\mu(x, y)$ is the Möbius function for $[x, y] := \{z \in L \mid x \preceq z \preceq y\}$
- $f([y, \hat{1}])$ is the number of maximal chains in the interval $[y, \hat{1}]$.
- Brown defined, for each element $x \in L$, a **derangement number** d_x

$$d_x = \sum_{y \succeq x} \mu(x, y) f([y, \hat{1}]) \ .$$

Spectrum of the Transition Matrix

Theorem (3)

Let P be a rooted forest, M the transition matrix of the promotion graph, and $\overline{M} = M + (x_1 + x_2 + \cdots + x_n)\mathbb{1}$. Then

$$\det(\overline{M} - \lambda \mathbb{1}) = \prod_{\substack{S \subseteq [n] \\ S \text{ upper set in } P}} (\lambda - x_S)^{d_S},$$

where $x_S = \sum_{i \in S} x_i$ and d_S is the derangement number in the lattice L (by inclusion) of upper sets in P.

In other words, for each upper set $S \subseteq [n]$, there is an eigenvalue x_S with multiplicity d_S.

Corollary

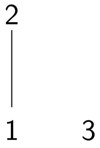
We consider the case of union of chains, and denote $P = [n_1] + [n_2] + \cdots + [n_k]$ to mean that the labels in the first chain are 1 through n_1, etc.

Lemma (4)

When P is a union of chains (labeled consecutively within chains), d_S is the number of linear extensions of $[n] \setminus S$ which are derangements.

Running Example

$P =$



- $\mathcal{L}(P) = \{123, 132, 312\}$
- Upper sets: ϕ, $\{2\}$, $\{3\}$, $\{2,3\}$, $\{1,2\}$, $\{1,2,3\}$
- Eigenvalues of $\overline{M}$: $x_1 + x_2 + x_3$, x_2, 0.

Another Example: The Tsetlin Library

- P is the n-antichain.
- $\mathcal{L}(P) = S_n$.
- Theorems 1 and 2 prove the formula about the stationary distribution.
- Theorem 3 and Lemma 4 proves the formula about the eigenvalues.

Definitions

- A **monoid** $\mathcal{M}$ is a set with an associative product and an identity.
- Natural **preorders** on $\mathcal{M}$:

$$x \leq_R y \text{ if } y = xu \text{ for some } u \in \mathcal{M}$$
$$x \leq_L y \text{ if } y = ux \text{ for some } u \in \mathcal{M}$$

- Equivalence classes on $\mathcal{M}$:

$$xRy \text{ if } y\mathcal{M} = x\mathcal{M}$$
$$xLy \text{ if } \mathcal{M}y = \mathcal{M}x$$

- $\mathcal{M}$ is R-trivial (L-trivial) if all R-classes (L-classes) are singletons. Equivalently, if the preorders are partial orders.

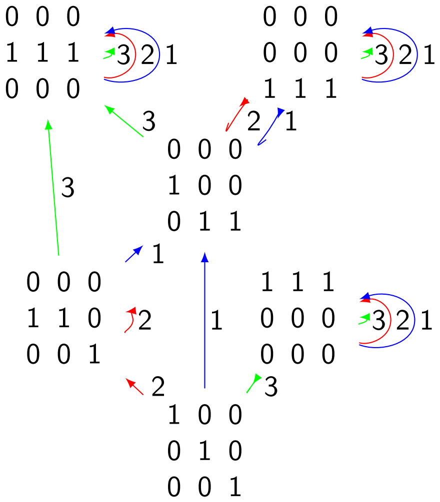
Promotion Monoid

- Define the operators G_i for $i \in [n]$ on $\mathcal{L}(P)$ by the promotion graph
$$G_i = \overline{M}_{x_1 = \cdots = x_{i-1} = x_{i+1} = \cdots = x_n = 0}.$$
- Define the **promotion monoid** $\mathcal{M}$ as the monoid generated by the G_i's.

Lemma (5)

$\mathcal{M}$ is R-trivial.

R-trivial monoid for the running example.



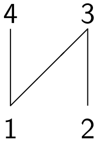
Proof ideas

- Construct the $\leq_R$ preorder on $\mathcal{M}$ and show that it is a partial order
- Prove an explicit eigenvalue formula for R-trivial monoids in general. This borrows ideas from Brown. Steinberg has such results for more general classes. This results in Theorem 3.
- Use that formula in Lemma 5 to describe degeneracies in terms of derangements. This proves Lemma 4.

Covering Times

- Recall that the uniform promotion graph led to the uniform distribution on linear extensions
- Counting linear extensions in an important problem in practice.
- Can we get better bounds on cover times? Or mixing times?
- One reason this is plausible is because the Markov chains are irreversible.

More classes of posets

- In general, the characteristic polynomial of the transition matrix does not factorize.
- But there are other classes for which it does.
- Eg: $P =$ 

$$0,\ -x_1-x_2,\ -x_1-x_2-x_4,\ -x_1-x_2-x_3-x_4,\ -2x_1-x_2-x_3-x_4.$$

Thank you for your attention!