Markov Chain Convergence in Discrete and Continuous Spaces

Prateek Bhakta

Markov chains are fundamental algorithmic tools used throughout the sciences and engineering, with applications ranging from combinatorial optimization, sampling and approximate counting, and evaluating the thermodynamic properties of physical systems. Their design and analysis has been a major focus of theoretical computer science for the last 20 years. Despite major advances, there are still fundamental questions that have resisted analysis and many new directions to pursue.

A Markov Chain is a probabilistic algorithm that takes a random walk in a large state space Ω that converges to a target stationary distribution π over Ω. The time required so that samples are chosen close to π is the *mixing time*. For the algorithm to be useful, the mixing time should be polynomial in the input to the problem, typically logarithmic in $|\Omega|$. Such a chain is called *rapidly mixing*.

I've been exploring interesting open questions in this field. The first is sampling weighted permutations, a simple discrete problem that has resisted analysis despite much attention. The second is a less studied direction extending known tools from discrete settings to continuous spaces.

1 Sampling Weighted Permutations

The first model was motivated by caching algorithms such as Move-Ahead-One and is a fundamental probabilistic model that has surprisingly resisted analysis. We consider a Markov chain on permutations of n integers, represented as a list, that randomly swaps adjacent entries. The Markov chain favors moves that place lower numbered entries in front of higher entries, and thus tends towards a sorted list.

More formally, we consider the following Markov chain on the space Ω of permutations of $\{1, \ldots, n\}$. We are given $p_{i,j}$ for $i, j \in \{1, \ldots, n\}$ for $i \neq j$ such that for $i < j$, $p_{i,j} \geq \frac{1}{2}$ and $p_{i,j} + p_{j,i} = 1$. The Markov chain M_{SWP} starts at any initial permutation σ. It then iteratively chooses a position $i \in 1, \ldots, n-1$ uniformly, and swaps the elements $\sigma(i), \sigma(i+1)$ with probability $p_{\sigma(i+1),\sigma(i)}$; else it does nothing.

Conjecture 1.1: The Markov Chain M_{SWP} is rapidly mixing for all $p_{i,j} \geq 1/2$ when $i < j$.

Jim Fill posed a related conjecture over 10 years ago under an assumption of convexity for the $\{p_{i,j}\}$. The convexity condition requires that $p_{i,j} \leq p_{i,j+1}$ for $1 \leq i < j \leq n-1$ and $p_{i,j} \leq p_{i-1,j}$ for $2 \leq i < j \leq n$.

Conjecture 1.2 (Fill): With the convexity condition, the spectral gap of the transition matrix of M_{SWP} is minimized when all $p_{i,j} = 1/2$ for $i \neq j$.

We are interested not in the spectral gap itself, but the related question of whether or not the chain is rapidly mixing. Note that both conjectures do not imply each other, but we believe that both are true.

Previous work: Despite the simplicity of the model presented, there are only a few cases for which the chain is known to be rapidly mixing. Benjamini et al. [1] proved that the above chain is rapidly mixing if for every $i < j$, $p_{i,j} = p*$ for any constant $\frac{1}{2} \leq p* \leq 1$. Greenburg, Pascoe, and Randall [4] have an alternate, simpler proof of this result using path coupling with a clever distance metric. Bubley and Dyer [2] showed that a similar Markov chain that samples linear extensions to a partial order is rapidly mixing, which implies that M_{SWP} is rapidly mixing if each $p_{i,j} \in \{1/2, 1\}$. Miracle, Pascoe, and Randall [7] considered the model where each integer $i \in \{1, \ldots, n\}$ is given a weight w_i such that $w_i \leq w_{i+1}$ for all $i < n$. We then let $p_{i,j} = \frac{w_j}{w_i + w_j}$ for all $i \neq j$. Under this model, they used a decomposition argument to show rapid mixing in cases where the integers take on two distinct weights. They also have rapid mixing results for a three-valued weight case, provided that two of the weights are very close to each other.

Our Approaches: My research explores conditions under which the above Markov chain is rapidly mixing. While attempting to solve the problem in its most general form, my advisor, Dana Randall, and I have identified some promising approaches and interesting special cases.

One compelling approach is to try to infer the mixing time of the chain on a given set of $\{p_{i,j}\}$ from the known mixing times of two closely related sets of $\{p_{i,j}\}$. For instance, fix $i \neq j \in \{1, \ldots, n\}$ and suppose we know that the chain is rapidly mixing when $p_{i,j} = 1/2$ and when $p_{i,j} = 1$, with all other $\{p_{k,l}\}$ fixed. Can we infer that the chain is rapidly mixing for all $1/2 < p_{i,j} < 1$?

Our goal then is to start with known cases where all $p_{i,j}$ are either 1/2 or 1 and use these arguments recursively, changing the value of one $p_{i,j}$ at a time, ensuring that there is no exponential increase of the mixing time. We know that when all $p_{i,j}$ are 1/2 or 1 that the Markov chain is rapidly mixing[2]. We have several results already:

- Our first approach is to bound the conductance. For any $S \subseteq \Omega$, let $\phi_S = \frac{\sum_{x \in S, y \in \bar{S}} \pi(x) P(x,y)}{\pi(S)}$. The conductance $\phi = min_{S \subseteq \Omega, \pi(S) \leq 1/2} \phi_S$ is a well known measure of the mixing rate. We have shown that for any i, j and every cut S, ϕ_S changes monotonically as a function of $p_{i,j}$ between $[1/2, 1]$. However, this does not solve the problem, as if $\pi(S) < 1/2$ on one endpoint of $[1/2, 1]$ and $\pi(S) > 1/2$ at the other endpoint, then we must "flip" the relevant half of the state space to measure conductance when $\pi(S) = 1/2$.

  Our technique has not yet exploited the connectivity structure of Ω, and deals with arbitrary cuts instead of those with minimum conductance. Thus one avenue of research will be to identify the minimum conductance cuts and to leverage the combinatorial structure of Ω to show that the min conductance ϕ in fact also changes monotonically.

- The maximum congestion of an edge in a set of canonical paths between each pair of vertices is also related to conductance and the mixing rate.[8]

  Assumption 1.1: There exists a set of canonical paths that exhibits small max congestion for cases when all $p_{i,j} = 1/2$ or $p_{i,j} = 1$, such that the canonical paths are also shortest paths (they do not flip any two elements i, j more than once).

  Under this assumption, we have shown that the congestion through any edge that does not transpose i, j changes monotonically as $p_{i,j}$ moves over $[1/2, 1]$, and the congestion of any edge that does transpose i, j will increase by at most a factor of 2. In the recursive process described above, every pair i, j is only changed once and thus the max congestion is at most twice the max congestion over all cases where $p_{i,j}$ is 1/2 or 1. This yields the desired result. However, it remains to find these canonical paths.

2 Bounding Mixing Times in Continuous Spaces

A second project that I am working on in collaboration with Dana is extending important tools for Markov chains from the discrete to the continuous setting. Notable examples where this has been accomplished include estimating the volume of a convex body [3][6], and mixing points on a line [10], and on a circle [9]. However, there are surprisingly few other examples where we have good bounds on mixing time. We are considering several models that have both a discrete and a continuous component in order to derive good bounds on mixing in continuous spaces.

In particular, we have been studying formulations of the well studied Ising model. The classical Ising model consists of a lattice of nodes, each of which contains a $\sigma_i = +1/-1$ "spin". As we vary the "temperature" parameter of the model, Jerrum and Sinclair [5] showed that the well studied Glauber Dynamics, or local moves Markov chain, exhibits a discrete phase transition, where the mixing time of the Markov chain jumps from polynomial to exponential time. It is one of the few chains were the exact location of the transition is known, and is believed to be intimately connected to the thermodynamic properties of magnets. I am working on a generalized version of this model that handles multiple spin orientations per vertex; we can imagine that there is a 2D unit vector at each point, pointing in n possible directions.

Our Approach: I have established bounds on the mixing rates at various temperatures for general n. One interesting special case is $n = 4$, where I've been able to prove another exact phase transition *at exactly the same temperature* as the classic Ising Model, where n=2. The technique decomposes the Markov Chain when $n = 4$ into two *independent* classic Ising models. Thus the mixing time of the case when $n = 4$ is roughly twice the mixing time of the classic Ising model, and thus both are polynomial or exponential in n concurrently. Though the analysis on higher n is not yet complete, there is some hope that exact transitions can be discovered for all n, possibly (though unlikely) all at the same temperature.

By taking $n \to \infty$, we can hope to study a continuous analogue of the Ising model, one that may provide more insight into the phase transitions of real world magnets, whose atoms can be oriented in a continuous space of configurations.

References

[1] Itai Benjamini, Noam Berger, and Christopher Hoffman. Mixing times of the biased card shuffling and the asymmetric exclusion process. *Trans. Amer. Math. Soc*, 2005.

[2] Russ Bubley and Martin Dyer. Faster random generation of linear extensions. In *Proceedings of the ninth annual ACM-SIAM symposium on Discrete algorithms*, SODA '98, 1998.

[3] Martin Dyer, Alan Frieze, and Ravi Kannan. A random polynomial-time algorithm for approximating the volume of convex bodies. *J. ACM*, 1991.

[4] Sam Greenberg, Amanda Pascoe, and Dana Randall. Sampling biased lattice configurations using exponential metrics. In *Proceedings of the twentieth Annual ACM-SIAM Symposium on Discrete Algorithms*, SODA '09, 2009.

[5] Mark Jerrum and Alastair Sinclair. Poly-time approximation algorithms for the ising model. In *Society for Industrial and Applied Mathematics*, 1993.

[6] László Lovász and Santosh Vempala. Simulated annealing in convex bodies and an $O(n^4)$ volume algorithm. In *Proceedings of the 44th Annual IEEE Symposium on Foundations of Computer Science*, FOCS '03, 2003.

[7] Sarah Miracle, Amanda Pascoe, and Dana Randall. Personal Communication, 2011.

[8] Dana Randall. Mixing. In *Proceedings of the 44th annual IEEE symposium on Foundations of Computer Science*, 2003.

[9] Dana Randall and Peter Winkler. Mixing points on a circle. In *9th International Workshop on Randomization and Approximation Techniques in Computer Science*, Random '05, 2005.

[10] Dana Randall and Peter Winkler. Mixing points on an interval. In *Proceedings of the Seventh Workshop on Algorithm Engineering and Experiments and the Second Workshop on Analytic Algorthmics and Combinatorics*, ALENEX/ANALCO '05, 2005.