

Nonexistence of Efficient Markovian Coupling for Finite State Markov Chains

Joseph Sweeney

University of Connecticut, 2018

ABSTRACT

Characterization of existence of an efficient Markovian coupling for a finite-state Markov chain is essentially an open problem. The goal of this research is to provide sufficient conditions on the transition function of a discrete-time finite-state Markov chain that guarantee nonexistence of an efficient Markovian coupling. Specifically, we look at symmetric, three state matrices with a constant antidiagonal. We show that an efficient Markovian coupling only exists when the matrix has a specific form.

Nonexistence of Efficient Markovian Coupling for Finite State Markov Chains

Joseph Sweeney

2018

Copyright by

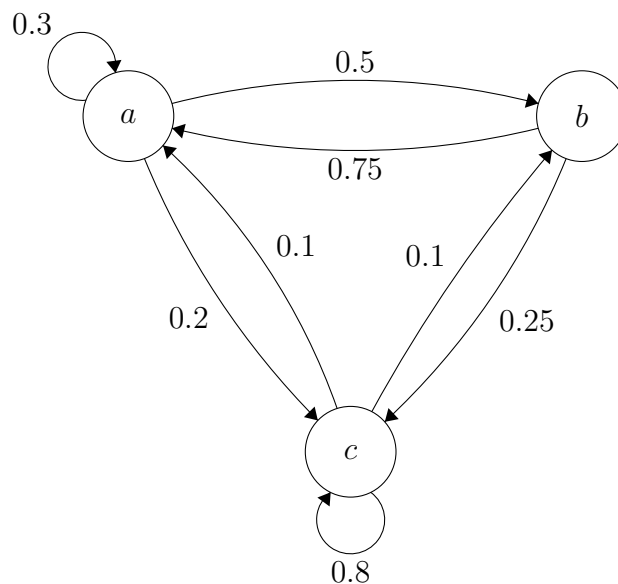
Joseph Sweeney

2018

Introduction

Background

This research has entirely been focussed around discrete time, finite state Markov Chains. A Markov Chain is a function of a single state and random values that output the next state. This can be thought of as the set of probabilities of moving from one state to any other. The key point here is that a Markov Chain only needs to know the current state, any information about prior states are ignored in these probabilities. The “discrete time” refers to the fact that there will be distinct steps at which we jump to the next state, in contrast to continuous time, where the change could happen at any moment. The “finite” or “countable” state refers to the fact that we have a countable number of states that we can be in. Here is a diagram that describes a Markov Chain on three states, a , b , and c .



Markov Chains where you are able move from one state to any other state given enough time will always have a unique stationary distribution as long as the proba-

bility of being at a state at time t converges to a distribution as $t \rightarrow \infty$. This is a probability distribution over every state in your space, and is stationary in the sense that if you apply the Markov Chain to this distribution, it remains the same. This distribution is also what the Markov Chain will converge to on the limit. In other words, if you start at any state and follow the Markov Chain long enough, keeping track of how many times you've visited each state, the final proportion of time spent in each state will be your stationary distribution.

Original Project

This research project was originally focussed on using the Metropolis Hastings algorithm to solve NP hard optimization projects. It started after learning that stochastic processes can be used to search a space probabilistically. This is done by doing a random search, where we set the probability of finding “high energy” states to be more likely.

The Metropolis Hastings algorithm allows us to find a Markov Chain that has any given stationary distribution. We can then define a distribution such that the probability of wanted states are higher than the probability of unwanted states. Then you just start in an arbitrary state, and follow the given Markov Chain which will eventually lead you to a desired state.

We tried to use this algorithm to tackle two-sided matching with couples. The problem is usually framed in terms of hospitals and residents, each with a ranked order of the opposite group. We want to create a stable matching between these two groups. This is a fairly straightforward problem with an $O(n^2)$ algorithm, but when you add the complication that some residents come in couples, so we need to ensure a

stable matching where all couples are geographically close to each other. The addition of couples actually makes this an NP hard problem.

Our solution attempted to frame this problem in terms of a probabilistic search on the space of all matchings, with a higher probability given to more stable matchings. Unfortunately, this approach was not very successful for a couple different reasons. First, the transition between matchings was too limited, each iteration we probabilistically switched the matching for two pairs, which meant we weren't able to search enough of the space to find stable matchings. Second, probabilistic search is excellent for problems where there is a gradual scale of usefulness for each state. This isn't the case for two-sided matching, where you only care about stable matchings, not almost stable matchings. After exploring this research for a few months we pivoted to a different question.

Mixing Time and Coupling

How can we put a bound on how long it takes to reach a stationary distribution for a given Markov Chain? In other words, starting from an arbitrary distribution, how many time steps will it take to get close to the stationary distribution? To approach this question, we need to introduce the idea of a *mixing time*. A mixing time is the minimum number of steps t , after which we are within some small ϵ of the stationary distribution. Formally, we define the *total variation distance* as $d_t(x, y)$

$$d_t(x, y) = \max_{A \subseteq S} (\mathbb{P}_x(X_t \in A) - \mathbb{P}_y(X_t \in A))$$

Where x and y are the starting states. The mixing time is the time t such that $d_t(x, y) \leq \epsilon$.

This mixing time can be difficult to find for an arbitrary Markov Chain. To help bound it, we will use a concept called a Markovian Coupling. This can be thought of as two copies of the Markov Chain starting at different states. We're going to look at the number of time steps that go by before they land on the same state which we'll call the *coupling time*. After they meet, they move together, never separating.

A coupling has two properties that make it appealing. The first is that it is generally easy to compute the coupling time. The second is that the coupling time can be used as an upper bound for the mixing time of a Markov Chain.

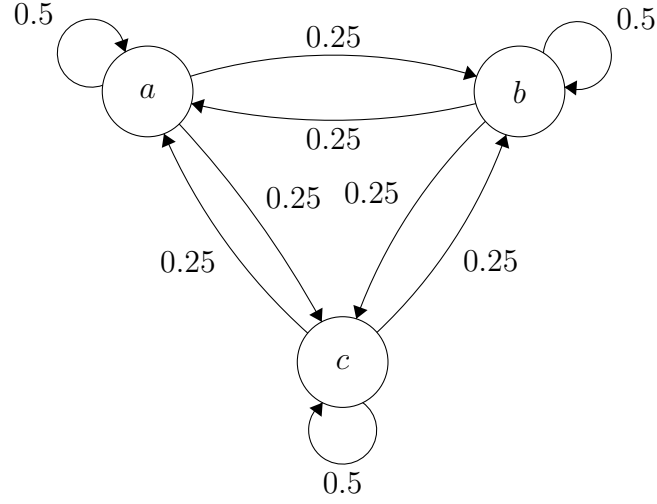
$$d_t(x, y) \leq \Pr_{x,y}(\tau_{couple} > t) \quad \forall t \in \mathbb{Z}_+$$

Where equality means you have a *maximal coupling*.

We'll also denote an *efficient coupling* as any coupling where there exists some constant c where

$$\Pr_{x,y}(\tau_{couple} > t) \leq c d_t(x, y) \quad \forall t \in \mathbb{Z}_+$$

As an example, let's say we have the following Markov Chain.



Our coupling can be as simple as two independent copies of this Markov Chain starting at points i and j . If we look at the probability of meeting at the next step we have

$$Pr(\text{meet}) = \sum_k p(i, k)p(j, k) = \frac{5}{16} = \alpha$$

Then the probability of not meeting at time t is

$$Pr(\text{not meet at } t) = (1 - \alpha)^t = \left(\frac{11}{16}\right)^t$$

Thus we know $d_t(x, y) \leq \left(\frac{11}{16}\right)^t$ for this Markov Chain.

There are other ways to create a coupling for this Markov Chain though, the main method we'll talk about is a greedy coupling. A greedy coupling is two copies of the same randomness that try to maximize the probability of meeting in the next step. For a greedy coupling on this chain, the probability of meeting at the next step is

$$Pr(\text{meet}) = \frac{3}{4}$$

because for any two states that you start on, the only way that you don't meet with the greedy coupling is if both copies remain in their states which has a probability of $\frac{1}{4}$. Then the probability of not meeting at time t is

$$Pr(\text{not meet at } t) = (0.25)^t$$

With $d_t(x, y) \leq (0.25)^t$, the greedy coupling gives us a tighter bound.

The greedy method of coupling can be expanded to optimize the chance of meeting in 2 steps, which we'll call the greedy 2-step Markovian Coupling.

Problem

The research project ended up focusing on trying to find general rules for when an efficient Markovian coupling exists when you couple using a purely greedy method. Specifically, we looked at reversible, 3 state Markov Chains and tried to find any properties of these Markov Chains that meant an efficient coupling was impossible.

To do this, we first wrote a simulation that would generate random 3 state Markov Chains, produce a greedy coupling, and then check whether or not this coupling is efficient. This made it possible to refine the motivating examples, so we can analytically solve the general case where we couldn't get an efficient coupling using the greedy method.

Results

In this section we show that for a large class of reversible 3-state chains there does not exist an efficient Markovian coupling.

Theorem 1. *Suppose that p is a symmetric transition function on three states with constant antidiagonal. Then p has an efficient Markovian coupling if and only if two entries in the first row of p are identical, and in this case the greedy coupling is efficient. Otherwise, any greedy 2-step Markovian coupling is inefficient.*

Proof. Let a, b represent the minimal element in the first row and the maximal element in the first row, respectively. Therefore the elements in the first row are (not necessarily in this order of appearance), $a \leq 1 - a - b \leq b$ and let $c = 1 - a - b$. Without loss of generality we can imagine the format of the matrix is

$$\begin{pmatrix} a & b & c \\ b & c & a \\ c & a & b \end{pmatrix}$$

By the assumption, the trace of p is equal to 1. Therefore, the eigenvalues for p are $1, \lambda, -\lambda$, and the determinant of p is equal to λ^2 . The trace of p^2 is therefore equal to $1 + 2\lambda^2$, but by symmetry of p and since the second and the third rows are the first row shifted, it follows that the diagonal of p^2 is constant, and its trace is equal to $3(a^2 + b^2 + (1 - a - b)^2)$. We therefore have

$$1 + 2\lambda^2 = 3(a^2 + b^2 + (1 - a - b)^2). \quad (1)$$

Suppose (X, Y) is a Markovian coupling for p and let τ be the coupling time. Then

$$\begin{aligned} P_{x,y}(\tau > t+1) &= \sum_{x',y'} P_{x,y}(\tau > t, X_t = x', Y_t = y', X_{t+1} \neq Y_{t+1}) \\ &= \sum_{x,y} P_{x,y}(X_{t+1} \neq Y_{t+1} | \tau > t, X_t = x', Y_t = y') P_{x,y}(\tau > t, X_t = x', Y_t = y'), \end{aligned} \tag{2}$$

where the summation is over (x', y') such that $P_{x,y}(\tau > t, X_t = x', Y_t = y') > 0$.

Clearly,

$$P_{x,y}(X_{t+1} \neq Y_{t+1} | \tau > t, X_t = x', Y_t = y') = 1 - \sum_{\ell} P_{x,y}(X_{t+1} = Y_{t+1} = \ell | \tau > t, X_t = x', Y_t = y').$$

Conditioned on $\tau > t, X_t = x', Y_t = y'$, the event $X_{t+1} = Y_{t+1} = \ell$ is contained both in events $X_{t+1} = \ell$ and $Y_{t+1} = \ell$. By the Markov property for each component process, the respective conditional probabilities of these events are $p(x', \ell)$ and $p(y', \ell)$, respectively. Note that we do not assume (X, Y) to be a Markov chain: we only assume each component is a Markov chain with respect to the joint filtration. Therefore

$$P_{x,y}(X_{t+1} = Y_{t+1} = \ell | \tau > t, X_t = x', Y_t = y') \leq \min(p(x', \ell), p(y', \ell)),$$

with equality if we choose a greedy coupling. From this we obtain

$$P_{x,y}(X_{t+1} \neq Y_{t+1} | \tau > t, X_t = x', Y_t = y') \geq 1 - \sum_{\ell} \min\{p(x', \ell), p(y', \ell)\},$$

In our particular case, for each x', y' the sum on the righthand side is equal to $2a + (1 - a - b) = a - b + 1$. Therefore we have proven that $P_{x,y}(\tau > t+1 | \tau > t, X_t =$

$x', Y_t = y') = b - a$, and plugging this equality in (2) and summing over x', y' , and a trivial induction then give

$$P_{x,y}(\tau > t + 1) \geq (b - a)^{t+1}, \quad \text{for all } t \in \mathbb{Z}_+, \quad (3)$$

with equality if the coupling is greedy. Since by Aldous' inequality $\lambda \leq b - a$ with equality if and only if the coupling is efficient, it follows from (1) that

$$f(a, b) := 1 + 2(b - a)^2 - 3(a^2 + b^2 + (1 - a - b)^2) \geq 0,$$

with equality if and only if the coupling is efficient. With $c = 1 - a - b$, and recall that $0 \leq a \leq c \leq b$. We can then simplify.

$$\begin{aligned} 0 &\leq 1 + (2a^2 + 2b^2 - 4ab) - 3a^2 - 3b^2 - 3c^2 \\ &\leq (1 - (a + b)^2) - 2ab - 3c^2 \\ &\leq (1 - a - b)(1 + a + b) - 2ab - 3c^2 \\ &\leq c(2 - 4c) - 2ab \\ &\leq c(1 - c - c) - ab \\ &\leq -(c^2 - c(a + b) + ab) \\ &\leq (b - c)(c - a) \end{aligned}$$

Therefore f is equal to zero if and only if $c = a$ or $c = b$. In this latter case the greedy coupling gives us an efficient coupling.

It remains to show that any 2-step greedy coupling is efficient. Let r_1, r_2, r_3 be the first second and third rows of p , respectively, written as row vectors. Then

$p^2(i, i) = r_1 \cdot r_1 = a^2 + b^2 + c^2$, and since the rows are cyclic permutations, and by Cauchy-Schwarz, it follows that $p(i, j) = r_1 \cdot r_2 \leq p(i, i)$ for all $i \neq j$. Therefore

$$p^2 = \begin{pmatrix} \beta & \alpha & \alpha \\ \alpha & \beta & \alpha \\ \alpha & \alpha & \beta \end{pmatrix}.$$

Repeating the argument leading to (3) verbatim, it follows that for any greedy coupling for p^2 , with coupling time σ , we have

$$P_{x,y}(\sigma > t) = (\beta - \alpha)^t.$$

To complete the proof, it is enough to show that $\beta - \alpha = \lambda^2$. However as the trace of p^2 is 3β and is also equal to $1 + 2\lambda^2$, we need to show $1 + 2(\beta - \alpha) = 3\beta$ always holds.

$$1 = 3\beta - 2(\beta - \alpha)$$

$$1 = 3\beta - 2\beta + 2\alpha$$

And this is equivalent to $2\alpha + \beta = 1$, which must always hold. □

Conclusion

To continue this research, the next steps would be to try and generalize this to any 3 state reversible Markov Chain. In addition, any generalizations to all 3 state or greater Markov Chains would push the boundary of knowledge in the field.

The general process of sampling random Markov Chains, and computing a cou-

pling could allow for further insight into questions of existence of efficient couplings for matrices that do not follow the criteria explored in this paper.