Essential Mathematics for the Political and Social Research

JEFF GILL
Cambridge University Press

Lecture Slides, Chapter 9: Markov Chains

Chapter 9 Objectives

- Studying Markov chains is also a nice reinforcement of important principles we have already covered.
- Markov chains are relevant to the things we study in the social sciences because they condition on the current status of events to say something about future events.
- Therefore they model how social and physical phenomena move from one state to another.
- Researchers find that this is a nice way to describe individual human decision making and collective human determinations.

As Time Goes By

- Suppose that your decision-making process is based only on a current state of affairs.
- Values at a previous point in time are irrelevant to future behavior.
- So stock purchase decisions, military strategy, travel directions, and other such trajectories can often successfully be described with Markov chains.

Technical Definition: Stochastic Process

- A *stochastic process* is a consecutive series of observed random variables.
- For some state space, Θ, the random variable θ is defined by:

$$\theta^{[t]} \sim F(\theta), \ t \in T,$$

  where t is some index value from the set T.
- It is typical to define T as the positive integers so $t = 0, 1, 2, 3, \ldots$.
- The state space that a stochastic process is defined on must be identified: it defines what values $\theta^{[t]}$ can take on any point in time t (discrete or continuous).

Technical Definition: Markov Chain

- A *Markov chain* is a stochastic process with the property that any specified state in the series, $\theta^{[t]}$, is dependent only on the previous state, $\theta^{[t-1]}$.
- Also the previous value, $\theta^{[t-1]}$ is then also conditional on its previous value, $\theta^{[t-2]}$, and so on.
- But "decisions" to move on the next future step only use information about the current location/value.
- This "memoryless" property can be explicitly stated in formal probability language: $\theta^{[t]}$ is *conditionally independent* on all values previous to $\theta^{[t-1]}$ if

$$p(\theta^{[t]} \in A | \theta^{[0]}, \theta^{[1]}, \ldots, \theta^{[t-2]}, \theta^{[t-1]}) = p(\theta^{[t]} \in A | \theta^{[t-1]}),$$

where A is any identified set (an event or range of events) on the complete state space.

Technical Definition: Martingale

- A different type of stochastic process that sometimes gets mentioned in the same texts is a *martingale*.

- A martingale is defined using expectation instead of probability:

$$E(\theta^{[t]} \in A|\theta^{[0]}, \theta^{[1]}, \ldots, \theta^{[t-2]}, \theta^{[t-1]}) = \theta^{[t-1]}.$$

- The expected value that the martingale is in the set A in the next period is the value of the current position.

- This differs from the Markov chain in that there is a stable iterative process based on this expectation rather than on Markovian probabilistic exploration.

As Time Goes By

- Since the future at time $t+1$ and the past at time $t-1$ are independent given the state at time t, the Markovian property does not care about the direction of time.
- This works because time here is not a physical characteristic of the universe; instead it is a series of our own construction. Interestingly, there are
- There are actually Markov chains that are defined to work backward through "time," such as those in *coupling from the past* (see Propp and Wilson 1996).

Being Discrete

- The discussion is restricted here to discrete-time, homogeneous Markov chains.
- By *discrete time*, we simply mean that the counting process above, $t = 0, 1, \ldots, T$, is recordable at clear, distinguishable points.
- Continuous-time Markov processes are substantially more abstract and we will not worry about them here.
- A *homogeneous* discrete Markov chain is one in which the process of moving (i.e., the *probability* of moving) is independent of current time: move decisions at time t are independent of t.

Contraception Use in Barbados

- Ebanks (1970) looked at contraception use by women of lower socio-economic class in Barbados and found a stable pattern in the 1950s and a different stable pattern emerged in the late 1960s.
- This is of anthropological interest because contraception and reproduction are key components of family and social life in rural areas.
- His focus was on the stability and change of usage, looking at a sample from family planning programs at the time.
- Using 405 respondents from 1955 and another 405 respondents from 1967, he produced the following change probabilities where the row indicates current state and the column indicates the next state (so, for instance, the probability of moving from "Use" at the current state to "Not Use" in the next state for 1955 is 0.52):

1955	Use	Not Use
Use	0.48	0.52
Not Use	0.08	0.92

1967	Use	Not Use
Use	0.89	0.11
Not Use	0.52	0.48

Contraception Use in Barbados

- Suppose we were interested in predicting contraception usage for 1969, that is, two years into the future.
- This could be done simply by the following steps:

$$\begin{bmatrix} 0.89 & 0.11 \\ 0.52 & 0.48 \end{bmatrix} \times \begin{bmatrix} 0.89 & 0.11 \\ 0.52 & 0.48 \end{bmatrix} = \underbrace{\begin{bmatrix} 0.85 & 0.15 \\ 0.71 & 0.29 \end{bmatrix}}_{1968}$$

$$\begin{bmatrix} 0.85 & 0.15 \\ 0.71 & 0.29 \end{bmatrix} \times \begin{bmatrix} 0.89 & 0.11 \\ 0.52 & 0.48 \end{bmatrix} = \underbrace{\begin{bmatrix} 0.83 & 0.17 \\ 0.78 & 0.22 \end{bmatrix}}_{1969}.$$

- This means that we would *expect* to see an increase in nonusers converting to users if the 1967 rate is an underlying trend.

Contraception Use in Barbados

- We can also test Ebanks' assertion that the 1950s were stable.
- Suppose we take the 1955 matrix of transitions and apply it iteratively to get a predicted distribution across the four cells for 1960.
- We can then compare it to the actual distribution seen in 1960, and if it is similar, then the claim is supportable (the match will not be exact, of course, due to sampling considerations).
- Multiplying the 1955 matrix four times gives

$$\begin{bmatrix} 0.48 & 0.52 \\ 0.08 & 0.92 \end{bmatrix}^4 = \begin{bmatrix} 0.16 & 0.84 \\ 0.13 & 0.87 \end{bmatrix}.$$

Contraception Use in Barbados

- This suggests the following empirical distribution, given the marginal numbers of users for 1959 in the study:

1959-1960 predicted	Use	Not Use
Use	7	41
Not Use	46	311

- This can be compared with the actual 1960 numbers from that study:

1959-1960 actual	Use	Not Use
Use	27	21
Not Use	39	318

- These are clearly dissimilar enough to suggest that the process is not Markovian as claimed.

The Markov Chain Kernel

- Recall that that a Markov chain moves based only on its current position.
- But using that information, how does the Markov chain decide?
- Every Markov chain is defined by two things: its state space (already discussed) and its *transition kernel*, $K()$.
- The *transition kernel* is a general mechanism for describing the probability of moving to other states based on the current chain status.
- Notationaly, $K(\theta, A)$ is a defined probability measure for all θ points in the state space to the set $A \in \Theta$: It maps potential transition events to their probability of occurrence.

Matrix Mapping

- The easiest case to understand is when the state space is discrete and $\mathbf{K}$ is just a matrix mapping: a $k \times k$ matrix for k discrete elements in that exhaust the allowable state space, A.
- Use the notation θ_i, meaning the ith state of the space.
- So a Markov chain that occupies subspace i at time t is designated $\theta_i^{[t]}$.
- Each individual cell defines the probability of a state transition from the first term to all possible states:

$$\mathbf{K} = \begin{bmatrix} p(\theta_1,\theta_1) & p(\theta_1,\theta_2) & \dots & p(\theta_1,\theta_{k-1}) & p(\theta_1,\theta_k) \\ p(\theta_2,\theta_1) & p(\theta_2,\theta_2) & \dots & p(\theta_2,\theta_{k-1}) & p(\theta_2,\theta_k) \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ p(\theta_{k-1},\theta_1) & p(\theta_{k-1},\theta_2) & & p(\theta_{k-1},\theta_{k-1}) & p(\theta_{k-1},\theta_k) \\ p(\theta_k,\theta_1) & p(\theta_k,\theta_2) & \dots & p(\theta_1,\theta_{k-1}) & p(\theta_k,\theta_k) \end{bmatrix}.$$

Matrix Mapping

- The first term in $p()$, constant across rows, indicates where the chain is at the current period and the column indicates potential destinations.
- Each matrix element is a well-behaved probability, $p(\theta_i, \theta_j) \geq 0, \ \forall i, j \in A$.
- The notation here can be a little bit confusing as it looks like a *joint distribution*, which it is not.
- The rows of $\mathbf{K}$ sum to one and define a conditional PMF because they are all specified for the same starting value and cover each possible destination in the state space.
- If we multiply the transition matrix (kernel) by itself j times, the result, $\mathbf{K}^j$, gives the *j-step transition matrix* for this Markov chain.

Example: Campaign Contributions

- Individuals who have contributed to a Congress member's campaign in the past are more likely than others to contribute in the next campaign cycle than others.
- This is why politicians keep and value donor lists, even including those who have given only small amounts in the past.
- Suppose that 25% of those contributing in the past to a given member are likely to do so again and only 3% of those not giving in the past are likely to do so now.
- The resulting transition matrix is denoted as follows:

$$\begin{array}{c} \\ \text{last period} \left\{ \begin{array}{c} \theta_1 \\ \theta_2 \end{array} \right. \end{array} \begin{array}{c} \overbrace{\begin{array}{cc} \theta_1 & \theta_2 \end{array}}^{\text{current period}} \\ \begin{bmatrix} 0.97 & 0.03 \\ 0.75 & 0.25 \end{bmatrix} \end{array},$$

where θ_1 is the state for no contribution and θ_2 denotes a contribution.

Example: Campaign Contributions

- If we start with a list of 100 names where 50 of them contributed last period and 50 did not, what number can we expect to have contribute from this list?

- In Markov chain language this is called a starting point or starting vector:

$$S_0 = \begin{bmatrix} 50 & 50 \end{bmatrix};$$

  that is, before running the Markov chain, half of the group falls in each category.

- To get to the Markov chain first state, multiply the initial state by the transition matrix:

$$S_1 = \begin{bmatrix} 50 & 50 \end{bmatrix} \begin{bmatrix} 0.97 & 0.03 \\ 0.75 & 0.25 \end{bmatrix} = \begin{bmatrix} 86 & 14 \end{bmatrix} = S_1.$$

- We would expect to get contributions from 14 off of this list.

Example: Campaign Contributions

- Since incumbent members of Congress enjoy a repeated electoral advantage for a number of reasons, assume that our member runs more consecutive races (and wins!).
- If we keep track of this particular list over time, what happens to our expected number of contributors?
- Move the Markov chain forward in time to find:

$$\text{Second state:} \quad S_2 = \begin{bmatrix} 86 & 14 \end{bmatrix} \begin{bmatrix} 0.97 & 0.03 \\ 0.75 & 0.25 \end{bmatrix} = \begin{bmatrix} 94 & 6 \end{bmatrix}$$

$$\text{Third state:} \quad S_3 = \begin{bmatrix} 94 & 6 \end{bmatrix} \begin{bmatrix} 0.97 & 0.03 \\ 0.75 & 0.25 \end{bmatrix} = \begin{bmatrix} 96 & 4 \end{bmatrix}$$

$$\text{Fourth state:} \quad S_4 = \begin{bmatrix} 96 & 4 \end{bmatrix} \begin{bmatrix} 0.97 & 0.03 \\ 0.75 & 0.25 \end{bmatrix} = \begin{bmatrix} 96 & 4 \end{bmatrix}.$$

- No matter how many times we run this chain forward from this point, the returned state will always be $[96, 4]$.

The Stationary Distribution of a Markov Chain

- Markov chains can have a *stationary distribution*: a distribution reached from iterating the chain until some point in the future where all movement probabilities are governed by a single probabilistic statement, regardless of time or position.
- This is equivalent to saying that when a Markov chain has reached its stationary distribution there is a single marginal distribution rather than the conditional distributions in the transition kernel.
- Define $\pi(\theta)$ as the stationary distribution of the Markov chain for θ on the state space A.
- $p(\theta_i, \theta_j)$ is the probability that the chain will move from θ_i to θ_j at some arbitrary step t, and $\pi^t(\theta)$ is the corresponding marginal distribution.
- The stationary distribution satisfies

$$\sum_{\theta_i} \pi^t(\theta_i) p(\theta_i, \theta_j) = \pi^{t+1}(\theta_j).$$

Shuffling Cards

- Use a Markov chain algorithm to shuffle a deck of cards such that the marginal distribution is *uniform*.
- So the objective (stationary distribution) is a uniformly random distribution in the deck: The probability of any one card occupying any one position is $1/52$.

Shuffling Cards

- Algorithm: take the top card and insert it uniformly randomly at some other point in the deck, and continue.
- Now simplify the problem (without loss of generality) to consideration of a deck of only three cards numbered $1, 2, 3$.
- The sample space for this setup is then given by

$$A = \{[1, 2, 3], [1, 3, 2], [2, 1, 3], [2, 3, 1], [3, 1, 2], [3, 2, 1]\},$$

  which has $3! = 6$ elements.
- A sample chain trajectory looks like

$$\begin{gathered} [1,3, 2] \\ [3,1, 2] \\ [1,3, 2] \\ [3,2, 1] \\ \vdots \end{gathered}$$

Shuffling Cards

- For example:

Action	Outcome	Probability
return to top of deck	$[3, 1, 2]$	$\frac{1}{3}$
put in middle position	$[1, 3, 2]$	$\frac{1}{3}$
put in bottom position	$[1, 2, 3]$	$\frac{1}{3}$

- To establish the potential outcomes we only need to know the current position of the deck and the probability structure (the kernel).

- Being aware of the position of the deck at time $t = 2$ tells us everything we need to know about the deck, and having this information means that knowing that the position of the deck at time $t = 1$ was $[1, 3, 2]$ is irrelevant to calculating the potential outcomes and their probabilities in the table above.

- Once the current position is established in the Markov chain, decisions about where to go are conditionally independent of the past.

Shuffling Cards

- Not every position of the deck is immediately reachable from every other position.
- For instance, the move from $[1, 3, 2]$ to $[3, 2, 1]$ is impossible because it would require at least one additional step.
- The transition kernel assigns positive (uniform) probability from each state to each reachable state in one step and zero probability to all other states:

$$K = \begin{bmatrix} 1/3 & 0 & 1/3 & 1/3 & 0 & 0 \\ 0 & 1/3 & 0 & 0 & 1/3 & 1/3 \\ 1/3 & 1/3 & 1/3 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1/3 & 1/3 & 1/3 \\ 1/3 & 1/3 & 0 & 0 & 1/3 & 0 \\ 0 & 0 & 1/3 & 1/3 & 0 & 1/3 \end{bmatrix}.$$

- Begin with a starting point at $[1, 2, 3]$ and look at the marginal distribution after each application of the transition kernel.
- Do this by pre-multiplying the transition kernel matrix by $[1, 0, 0, 0, 0, 0]$ as the starting point.
- Then we record the result, multiply it by the kernel, and continue.

Shuffling Cards

- The first 15 iterations produce the following marginal probability vectors:

Iteration	p([1,2,3])	p([1,3,2])	p([2,1,3])	p([2,3,1])	p([3,1,2])	p([3,2,1])
1	0.222222	0.111111	0.222222	0.222222	0.111111	0.111111
2	0.185185	0.148148	0.185185	0.185185	0.148148	0.148148
3	0.172839	0.160493	0.172839	0.172839	0.160493	0.160493
4	0.168724	0.164609	0.168724	0.168724	0.164609	0.164609
5	0.167352	0.165980	0.167352	0.167352	0.165980	0.165980
6	0.166895	0.166438	0.166895	0.166895	0.166438	0.166438
7	0.166742	0.166590	0.166742	0.166742	0.166590	0.166590
8	0.166692	0.166641	0.166692	0.166692	0.166641	0.166641
9	0.166675	0.166658	0.166675	0.166675	0.166658	0.166658
10	0.166669	0.166663	0.166669	0.166669	0.166663	0.166663
11	0.166667	0.166665	0.166667	0.166667	0.166665	0.166665
12	0.166667	0.166666	0.166667	0.166667	0.166666	0.166666
13	0.166666	0.166666	0.166666	0.166666	0.166666	0.166666
14	0.166666	0.166666	0.166666	0.166666	0.166666	0.166666
15	0.166666	0.166666	0.166666	0.166666	0.166666	0.166666

Shuffling Cards

- The probability structure of the marginals have converged to a uniform pattern, as desired.
- These stationary probabilities are not the probabilities that govern movement at any particular point in the chain.
- These are the *unconditional* probabilities of seeing one of the six events at any arbitrary point in time, unconditional on current placement.
- Recall that there are three unavailable outcomes and three equal probability outcomes for each position of the deck.
- So the marginal distribution above *cannot* be functional as a state to state transition mechanism.
- In some future state it is equally likely that any of the six states would be observed, but this ignores the state of the deck at that point.

Properties of Markov Chains: Homogeneity

- A Markov chain is said to be homogeneous at some step t if the transition probabilities at this step do not depend on the value of t.
- This means that Markov chains can be homogeneous for some periods and *non-homogeneous* for other periods.
- The homogeneity property is usually important in that Markov chains that behave according to some function of their age are usually poor theoretical tools for exploring probability statements of interest.

Properties of Markov Chains: Periodicity

- If a Markov chain operates on a deterministic repeating schedule of steps, then it is said to be a Markov chain of *period-n*, where n is the time (i.e., the number of steps) in the reoccurring period.
- A periodic Markov chain is not a homogeneous Markov chain because the period implies a dependency of the chain on the time t.
- Markov chains are generally implemented with computers, and the underlying random numbers generated on computers have characteristics that make them not truly random, (thus called pseudo-random).
- Fortunately, the algorithms are sufficiently sophisticated that we can treat these values as truly random.

Properties of Markov Chains: Irreducibility

- A *state* A in the state space of a Markov chain is *irreducible* if for every two substates or individual events θ_i and θ_j in A, these two substates "communicate."
- This means that the Markov chain is irreducible on A if every reached point or collection of points can be reached from every other reached point or collection of points.:

$$p(\theta_i, \theta_j) \neq 0, \ \forall \theta_i, \theta_j \in A.$$

Properties of Markov Chains: Irreducibility

- As an example, consider the following kernel of a *reducible* Markov chain:

$$K = \begin{array}{c} \\ \theta_1 \\ \theta_2 \\ \theta_3 \\ \theta_4 \end{array} \begin{array}{c} \theta_1 \; \theta_2 \; \theta_3 \; \theta_4 \\ \begin{pmatrix} \frac{1}{2} & \frac{1}{2} & 0 & 0 \\ \frac{1}{3} & \frac{2}{3} & 0 & 0 \\ 0 & 0 & \frac{3}{4} & \frac{1}{4} \\ 0 & 0 & \frac{1}{4} & \frac{3}{4} \end{pmatrix} \end{array}.$$

- The chain determined by this kernel is reducible because if is started in either θ_1 or θ_2, then it operates as if has the transition matrix:

$$K_{1,2} = \begin{array}{c} \\ \theta_1 \\ \theta_2 \end{array} \begin{array}{c} \theta_1 \; \theta_2 \\ \begin{pmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{3} & \frac{2}{3} \end{pmatrix} \end{array},$$

- If it is started in either θ_3 or θ_4, then it operates as if it has the transition matrix:

$$K_{3,4} = \begin{array}{c} \\ \theta_3 \\ \theta_4 \end{array} \begin{array}{c} \theta_3 \; \theta_4 \\ \begin{pmatrix} \frac{3}{4} & \frac{1}{4} \\ \frac{1}{4} & \frac{3}{4} \end{pmatrix} \end{array}.$$

Properties of Markov Chains: Irreducibility

- The original Markov chain determined by K is *reducible* to one of two forms, depending on where it is started because there are in each case two permanently unavailable states.
- To provide a contrast, consider the Markov chain determined by the following kernel, K', which is very similar to K but is irreducible:

$$K = \begin{array}{c} \\ \theta_1 \\ \theta_2 \\ \theta_3 \\ \theta_4 \end{array} \begin{array}{c} \begin{array}{cccc} \theta_1 & \theta_2 & \theta_3 & \theta_4 \end{array} \\ \left(\begin{array}{cccc} \frac{1}{2} & \frac{1}{2} & 0 & 0 \\ \frac{1}{3} & 0 & \frac{2}{3} & 0 \\ 0 & \frac{3}{4} & 0 & \frac{1}{4} \\ 0 & 0 & \frac{1}{4} & \frac{3}{4} \end{array} \right) \end{array}.$$

- This occurs because there is now a two-way "path" between the previously separated upper and lower submatrices.

Properties of Markov Chains: Hitting Times

- The *hitting time* of a state A and a Markov chain θ is the shortest time for the Markov chain to begin in A and return to A:
$$T_A = \inf[n > 0, \theta^{[n]} \in A].$$
- Recall that the notation "inf" means the lowest (positive) value of n that satisfies $\theta^{[n]} \in A$.
- It is conventional to define T_A as infinity if the Markov chain never returns to A.

Properties of Markov Chains: Hitting Times

- From this definition we get the following important result:

  An irreducible and aperiodic Markov chain on state space A will for each subspace of A, a_i, have a finite hitting time for a_i with probability one and a finite expected value of the hitting time:

$$p(T_{a_i} < \infty) = 1, \qquad E[T_{a_i}] < \infty.$$

- By extension we can also say that the probability of transitioning from two substates of A, a_i and a_j, in finite time is guaranteed to be nonzero.

Properties of Markov Chains: Recurrence

- Some characteristics belong to *states* rather than Markov chains.
- Markov chains operating on these states are affected by such characteristics.
- A state A is said to be *absorbing* if once a Markov chain enters this state it cannot leave: $p(A, A^{\complement}) = 0$.
- Conversely, A is *transient* if the probability of the chain not returning to this state is nonzero: $1 - p(A, A) > 0$.
- This is equivalent to saying that the chain will return to A for only a finite number of visits in infinite time.
- State A is said to be *closed* to another state B if a Markov chain on A cannot reach B: $p(A, B) = 0$.
- State A is clearly closed *in general* if it is absorbing since $B = A^{\complement}$ in this case.

Properties of Markov Chains: Recurrence

- If a state is closed, discrete, and irreducible, then this state and all subspaces within this subspace are called *recurrent*, and Markov chains operating on recurrent state spaces are recurrent.
- From this we can say something important in two different ways:
  - An irreducible Markov chain is called *recurrent* with regard to a given state A, which is a single point or a defined collection of points,
  - if the probability that the chain occupies A infinitely often over unbounded time is nonzero.
- This means that when a chain moves into a recurrent state, it stays there forever and visits every subspace infinitely often.

Properties of Markov Chains: Recurrence

- There are two different, mutually exclusive, “flavors” of recurrence with regard to a state A:
  - A Markov chain is *positive recurrent* if the mean time to return to A is bounded.
  - Otherwise the mean time to return to A is infinite, and the Markov chain is called *null recurrent*.
- With these we can also state the following properties:
  - **Unions:** If A and B are recurrent states, then $A \cup B$ is a recurrent state.
  - **Capture:** A chain that enters a closed, irreducible, and recurrent state stays there and visits.

Conflict and Cooperation in Rural Andean Communities

- Robbins and Robbins (1979) extended Whyte's (1975) study of 12 Peruvian communities by extrapolating future probabilities of conflict and cooperation using a Markov chain analysis.

- Whyte classified these communities in 1964 and 1969 as having one of four types of relations with the other communities: high cooperation and high conflict (HcHx), high cooperation and low conflict (HcLx), low cooperation and high conflict (LcHx), or low cooperation and low conflict (LcLx).

- The interesting questions were, what patterns emerged as these communities changed (or not) over the five-year period since conflict and cooperation can exist simultaneously but not easily.

Conflict and Cooperation in Rural Andean Communities

- The states of these communities at the two points in time are:

Community	Type in 1964	Type in 1969
Huayopampa	HcLx	HcHx
Pacaraos	HcHx	HcHx
Aucallama	LcHx	LcLx
La Esperanza	LcLx	LcLx
Pucará	LcHx	LcLx
St. A De Caias	LcHx	LcHx
Mito	LcHx	LcLx
Virú	LcHx	LcLx
Pisaq	LcHx	LcHx
Kuyo Chico	HcLx	HcLx
Maska	HcLx	HcLx
Qotobamba	HcLx	HcLx

Conflict and Cooperation in Rural Andean Communities

- If we are willing extrapolate these changes as Robbins and Robbins did by assuming that "present trends continue," then a Markov chain transition matrix can be constructed from the empirically observed changes between 1964 and 1969.
- This is given the following matrix, where the rows indicate 1964 starting points and the columns are 1969 outcomes:

$$\begin{array}{c} \\ \text{HcHx} \\ \text{HcLx} \\ \text{LcHx} \\ \text{LcLx} \end{array} \begin{array}{c} \begin{array}{cccc} \text{HcHx} & \text{HcLx} & \text{LcHx} & \text{LcLx} \end{array} \\ \begin{pmatrix} 1.00 & 0.00 & 0.00 & 0.00 \\ 0.25 & 0.75 & 0.00 & 0.00 \\ 0.00 & 0.00 & 0.33 & 0.67 \\ 0.00 & 0.00 & 0.00 & 1.00 \end{pmatrix} \end{array}.$$

Conflict and Cooperation in Rural Andean Communities

- HcHx and LcLx are both absorbing states as described above: once the Markov chain reaches these states it never leaves.
- Clearly this means that the Markov chain is not irreducible because there are states that cannot "communicate."
- There are two noncommunicating state spaces given by the 2×2 upper left and lower right submatrices.
- Intuitively it seems that any community that starts out as HcHx or HcLx ends up as HcHx (upper left), and any community that starts out as LcHx or LcLx ends up as LcLx (lower right).

Conflict and Cooperation in Rural Andean Communities

- We can test this by running the Markov chain for some reasonable number of iterations and observing the limiting behavior.
- It takes about 25 iterations (i.e., 25 five-year periods under the assumptions since the 0.75 value is quite persistent) for this limiting behavior to converge to the stationary state:

$$\begin{array}{c} \text{HcHx} \\ \text{HcLx} \\ \text{LcHx} \\ \text{LcLx} \end{array} \overset{\begin{array}{cccc} \text{HcHx} & \text{HcLx} & \text{LcHx} & \text{LcLx} \end{array}}{\begin{pmatrix} 1.00 & 0.00 & 0.00 & 0.00 \\ 1.00 & 0.00 & 0.00 & 0.00 \\ 0.00 & 0.00 & 0.00 & 1.00 \\ 0.00 & 0.00 & 0.00 & 1.00 \end{pmatrix}},$$

Properties of Markov Chains: Stationarity and Ergodicity

- In many applications a stochastic process eventually converges to a single limiting value and stays at that value permanently.
- It should be clear that a Markov chain *cannot* do that because it will by definition continue to move about the parameter space.
- Instead we are interested in the *distribution* that the Markov chain will eventually settle into.
- Actually, these chains do not *have* to converge in distribution, and some Markov chains will wander endlessly without pattern or prediction.

Properties of Markov Chains: Stationarity and Ergodicity

- We want the marginal distribution of a Markov chain.

- For a Markov chain operating on a discrete state space, the marginal distribution of the chain at the m step is obtained by inserting the current value of the chain, $\theta_i^{[m]}$, into the row of the transition kernel for the mth step, p^m:

$$p^m(\theta) = [p^m(\theta_1), p^m(\theta_2), \ldots, p^m(\theta_k)].$$

- The marginal distribution at the very first step of the discrete Markov chain is given by $p^1(\theta) = p^1\pi^0(\theta)$, where p^0 is the initial starting value assigned to the chain and $p^1 = p$ is a transition matrix.

- The marginal distribution at some (possibly distant) step for a given starting value is

$$p^n = pp^{n-1} = p(pp^{n-2}) = p^2(pp^{n-3}) = \ldots = p^n p^0.$$

- This shows how Markov chains eventually "forget" their starting points.

Properties of Markov Chains: Stationarity and Ergodicity

- Recall that $p(\theta_i, \theta_j)$ is the probability that the chain will move from θ_i to θ_j at some arbitrary step t.
- $\pi^t(\theta)$ is the corresponding marginal distribution. Define $\pi(\theta)$ as the stationary distribution (a well-behaved probability function in the Kolmogorov sense) of the Markov chain for θ on the state space A, if it satisfies

$$\sum_{\theta_i} \pi^t(\theta_i) p(\theta_i, \theta_j) = \pi^{t+1}(\theta_j).$$

Properties of Markov Chains: Stationarity and Ergodicity

- This means that the marginal distribution remains fixed when the chain reaches the stationary distribution, and we can drop the superscript designation for iteration number and just use $\pi(\theta)$; in shorthand, $\pi = \pi p$.

- Once the chain reaches its stationary distribution, it stays in this distribution and moves about, or "mixes," throughout the subspace according to marginal distribution, $\pi(\theta)$, indefinitely.

- The key theorem is

  An irreducible and aperiodic Markov chain will eventually converge to a stationary distribution, and this stationary distribution is unique.

- Here the recurrence gives the range restriction property whereas stationarity gives the constancy of the probability structure that dictates movement.

Properties of Markov Chains: Stationarity and Ergodicity

- If a chain is recurrent and aperiodic, then we call it *ergodic*, and ergodic Markov chains with transition kernel K have the property

$$\lim_{n\to\infty} K^n(\theta_i, \theta_j) = \pi(\theta_j),$$

  for all θ_i and θ_j in the subspace.

- *Once an ergodic Markov chain reaches stationarity, the resulting values are all from the distribution* $\pi(\theta_i)$.

- The Ergodic Theorem is the equivalent of the strong law of large numbers but instead for Markov chains, since it states that any specified function of the posterior distribution can be estimated with samples from a Markov chain in its ergodic state because averages of sample values give strongly consistent parameter estimates.

Properties of Markov Chains: Stationarity and Ergodicity

- The ergodic theory states that, given the right conditions, we can collect empirical evidence from the Markov chain values in lieu of analytical calculations.

- Specifically

  If θ_n is a positive recurrent, irreducible Markov chain with stationary distribution given by $\pi(\theta)$, then

$$\lim_{n\to\infty} \frac{1}{n} \sum f(\theta_n) = \sum_{\Theta} f(\theta)\pi(\theta).$$

- This means that empirical averages for the function $f()$ converge to probabilistic averages.

Population Migration Within Malawi

- Discrete Markov chains are enormously useful in demographics for describing movements of populations.
- Segal (1985) looked at population movements between Malawi's three administrative regions from 1976 to 1977.
- The Republic of Malawi is a narrow, extended south African country of 45,745 square miles wrapped around the eastern and southern parts of Lake Malawi.
- Segal took observed migration numbers to create a transition matrix for future movements under the assumption of stability.
- This is given by

$$\text{Source} \quad \begin{array}{c} \\ \text{Northern} \\ \text{Central} \\ \text{Southern} \end{array} \overset{\text{Destination}}{\begin{array}{ccc} \text{Northern} & \text{Central} & \text{Southern} \end{array}} \begin{pmatrix} 0.970 & 0.019 & 0.012 \\ 0.005 & 0.983 & 0.012 \\ 0.004 & 0.014 & 0.982 \end{pmatrix}.$$

Population Migration Within Malawi

- The question is whether the transition matrix given defines an ergodic Markov chain.
- Since this is a discrete transition kernel, we need to assert that it is recurrent and aperiodic.
- Recurrence comes from the lack of zero probability values in the matrix.
- Although the presence of zero probability values alone would not be proof of nonreccurence, the lack of any shows that all states communicate with nonzero probabilities and thus recurrence is clear.
- There is also no mechanism to impose a cycling effect through the cell values so aperiodicity is also apparent.
- Therefore this transition kernel defines an ergodic Markov chain that must then have a unique stationary distribution.

Population Migration Within Malawi

- Long periods of unchanging marginal probabilities are typically a good sign, especially with such a simple and well-behaved chains.
- The resulting stationary distribution after multiplying the transition kernel 600 times is

$$\begin{array}{ccc} & \text{Destination} & \\ \text{Northern} & \text{Central} & \text{Southern} \\ \end{array}$$
$$\begin{pmatrix} 0.1315539 & 0.4728313 & 0.3956149 \end{pmatrix}.$$

- Running the chain much longer does not change the observed results.
- At the original transition matrix, there is a strong inclination to stay in the same region of Malawi for each of the three regions (the smallest has probability 0.97), yet in the limiting distribution there is a markedly different result, with migration to the Central Region being almost 50%.
- Even though there is a 0.97 probability of remaining in the Northern Region for those starting there on any given cycle, the long-run probability of remaining in the Northern Region is only 0.13.

Properties of Markov Chains: Reversibility

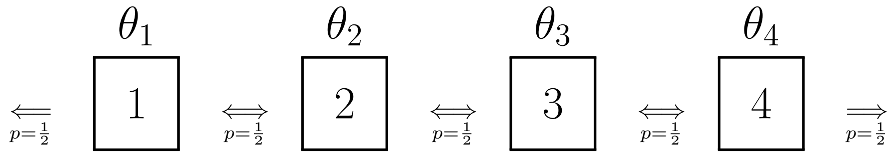
- Some Markov chains are *reversible* in that they perform the same run backward as forward.
- If $p(\theta_i, \theta_j)$ is a single probability from a transition kernel K and $\pi(\theta)$ is a marginal distribution, then the Markov chain is reversible if it meets the condition

$$p(\theta_i, \theta_j)\pi(\theta_i) = p(\theta_j, \theta_i)\pi(\theta_j).$$

- This expression is called both the *reversibility condition* and the *detailed balance equation*.
- It means is that the *distribution* of θ at time $t+1$ conditioned on the *value* of θ at time t is the same as the *distribution* of θ at time t conditioned on the *value* of θ at time $t+1$.
- Thus, for a reversible Markov chain the direction of time is irrelevant to its probability structure.

Properties of Markov Chains: Reversibility

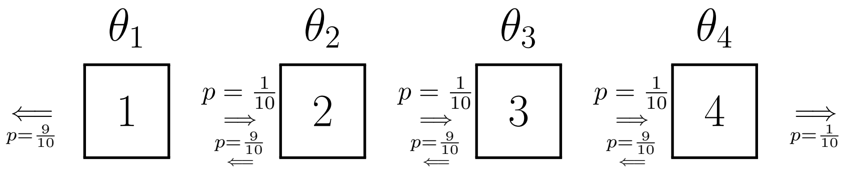
- As an example of reversibility, we modify a previous example where the probability of transitioning between adjacent states for a four-state system is determined by flipping a fair coin (states 1 and 4 are assumed adjacent to complete the system):



Properties of Markov Chains: Reversibility

- The stationary distribution of this system is uniform across the four events, and that it is guaranteed to reach it since it is recurrent and aperiodic.

- Suppose we modify the transition rule to be asymmetric from every point, according to:



Properties of Markov Chains: Reversibility

- This is a chain that strongly prefers to move left at every step by the same probability.
- This Markov chain will also lead to a uniform stationary distribution, because it is clearly still recurrent and aperiodic.
- It is not reversible anymore because for adjacent θ values (i.e., those with nonzero transition probabilities):

$$p(\theta_i, \theta_j)\pi(\theta_i) \neq p(\theta_j, \theta_i)\pi(\theta_j), \; i < j$$
$$\frac{1}{10}\frac{1}{4} \neq \frac{9}{10}\frac{1}{4}$$

(where we say that $4 < 1$ by assumption to complete the system).