# Chapter 5

# Markov Processes

*History is a cyclic poem written by time upon the memories of man.*

Percy Bysshe Shelley

The second nonlinear model that I'll describe are called Markov Processes. The name sounds kind of scary, but they're easy to understand. To frame my presentation of Markov Processes, I pick up on a point raised at the end of the previous chapter: the shortcomings of linear extrapolation. At present, we see a general trend toward more democratic countries. And one might well think that eventually, everyone will live in democracy. That might be pretty to think, but unfortunately given some rather mild assumptions the trend to democracy should slow, and in the long run, the entire world probably won't be democratic. Just like, in the long run, we won't all be living in California.

The democratization example provides an entree into thinking about linkages between *states of the world*, the key assumption in Markov theory. When a modeler mentions a state of the world she does not mean Kazakhstan or even Idaho. She means the relevant features of the world. So, the state of a country might be democratic or dictatorship. The state of a boat might be at dock, a sea and afloat, sinking, or sunk. Markov processes consist two parts: (1) a finite number of states and (2) transition probabilities for moving between those states. For example, on a given day, a person might be happy, sad, or feel a sense of ennui. These would be the states of the Markov process. The transition probabilities represent the likelihood of moving between those states – from a state of happiness on one day to a state of ennui the next.

Markov processes provide a powerful lens for viewing the world. Provided a small number of assumptions are met: a fixed set of states, fixed transition probabilities, and the possibility of getting from any state to another through a series of transitions, a Markov process converges to a unique distribution over states. This means that what happens in the long run won't depend on where the process started or on what happened along the way. What happens in the long run will be completely determined by the transition probabilities – the likelihoods of moving between the various states.

Endowed with knowledge of Markov Processes, we have a new way to think about the benefits of taking action. If a system follows a Markov Process, then initial conditions, interventions, and history itself have no bearing on the long run distribution over states. To put this in the context of an example, if your mental state can be characterized as a Markov Process, then any effort to move you from sadness to happiness will only have temporary effects. In the long run, you'll settle into a steady distribution over happy, sad, and ennui that's unaffected by bursts of joy. Similarly, if social processes are Markov Processes, then redistribution of resources would also have no long term effect.

When I teach Markov Processes, students often undergo a period of initial confusion as they think I am saying that history doesn't matter. I'm not saying that at all. What I am saying is that if events follows a Markov process, then history doesn't matter. So, If we want to believe (as I do by the way), that history does matter in the social world, then one of the assumptions of the Markov model must not hold. Thus, the model provides a framework to which we can compare the world. And it offers up a challenge: If the world fits the assumptions of the model, then what happens in the long run is set in stone. If we believe history matters, then it's incumbent upon us to show that at least one of the assumptions of the model doesn't hold. Otherwise, we're ignoring some pretty basic logic.

In what follows. I first show the data on democratization, which looks pretty compelling. I then describe a Markov Process that includes only two states and present some general theoretical results. I then go back to the data on democratization and view the trends through the lens of Markov theory. From the perspective of the Markov model, we should expect democratization to level out. I then lay describe the general Markov model. I conclude with some comments on the many uses of the Markov approach and also return to, and eventually leave, California.

## Democratic Vistas

First, the data on democratization: Freedom House categories countries as either free, partly free or not free. The figure below shows the percentage of countries in each category over the past thirty-five years. The figure shows a clear trend toward increased democratization.

It's easy to make a linear approximation of any one of these lines project into the future. In the past thirty-five years, the percentage of free countries has risen approximately twenty percent. If that trend were to continue, by 2045, two thirds of all countries will be free and by 2080, nearly eight out of every nine countries will be free. And by the turn of the next century, almost every country will be free.

This is a happy thought. Yet, a closer look at the data reveals that the transition go in both directions: some countries move from partly free to free and from not free to partly free, but some move from the free category to the partly free and some even move from partly free to not free. These movements back and forth between categories suggest that

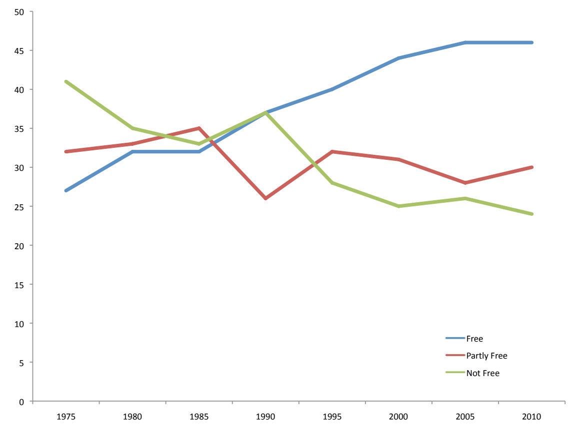


Figure 5.1: Freedom House Democratization Data

democratization shouldn't be seen as a simple linear trend. But how do we model it?

## Happy or Sad: a Two State Markov Process

Good modeling requires starting simple. I begin with simplest possible Markov Process. In a Markov Process, the thing being considered, be it an economy, a person, a machine, or an ecosystem, takes on one of some finite number of *states*. In the types of Markov Processes that I consider here time moves in discrete steps. These steps could be counted out in seconds, years, or days.[1] In each time step, the state of the system updates according to a *transition mapping*. I say that the state updates rather than say that it changes because it's possible that the state doesn't change. This transition mapping is typically probabilistic not deterministic, though nothing in what follows precludes deterministic transition mappings. What's important is that the transition probabilities stay fixed. They never change.

[1]In more general models this assumption can be relaxed. Throughout the book, I use the least technical versions of the models that still carry some weight.

The process I describe here has only two states. Each state describes a person's mental mood. I'll call these states *happy* and *sad*. On any given day, any person will be in exactly one of these states. Further, the person's state tomorrow will depend only on her state today.

Suppose that I've gathered a lot of data and found the following transitions between states. A person who is happy today has a ninety percent chance of being happy tomorrow and a ten percent chance of being sad. In contrast, an unhappy person today has only a thirty percent chance of being happy tomorrow and a seventy percent chance of being unhappy. Given these assumptions, if I start on a Monday with fifty happy people (represented by the yellow circle in the figure below) and fifty sad people (represented by the blue circle). On Tuesday, five of the happy people (ten percent of them) will become sad, and fifteen of the sad people (thirty percent) will become happy. This creates sixty happy people and forty sad people. Similar calculations yield that on Wednesday, sixty-six people are happy. On Thursday, approximately seventy people are happy. And on Friday about seventy-two people are happy.

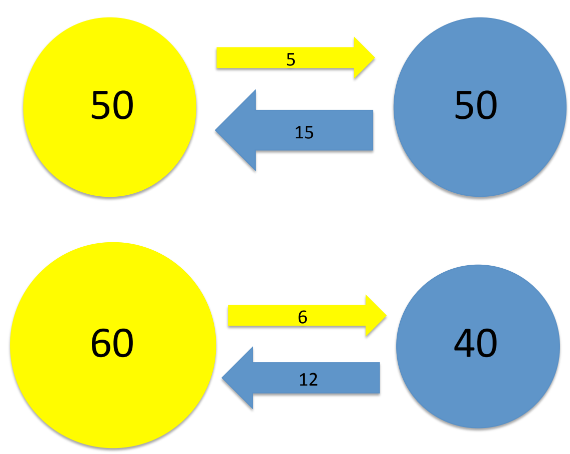


Figure 5.2: The first two steps of a Markov Process

Suppose instead, that I begin with one hundred happy people on Monday. On Tuesday, ninety people will be happy. On Wednesday, eighty-four will be happy. On Thursday, approximately eighty will be happy, and on Friday, seventy-eight will be happy. This number is not too far from the seventy-two that were happy when I began with equal numbers of happy and sad people.

In fact, if I were to continue to iterate through these two examples, I would find that the process converges on exactly seventy-five happy people and twenty-five unhappy people. This holds more generally. If I were to start with any number of happy people, any number from zero to one hundred, eventually, I'll have seventy-five happy people.

This distribution of seventy-five happy people and twenty-five unhappy people is called an *equilibrium*. Once the processes reaches those numbers, it stays there. To see why this is an equilibrium, we need only apply our transition probabilities as shown in the figure below. If seventy-five people are happy, then 7.5 will become unhappy next period. And, if twenty-five people are unhappy, then 7.5 will become happy. These two effects balance, so the result is an equilibrium.

As equilibria go, this one seems rather odd. Each of the hundred people moves back and forth between happy and sad. So, the system churns constantly. However, the expected number of people in each state doesn't change, so by convention, we say the system reaches a *statistical equilibrium*. This means that some statistic, in this case the number of happy people, reaches an equilibrium, even though the process keeps churning. In equilibrium, each day, seven and a half people move from happy to sad, and seven and a half people move from sad to happy. The process creates a cyclic poem.[2]

[2] If the half person concept bothers you, think of these as percentages of people.

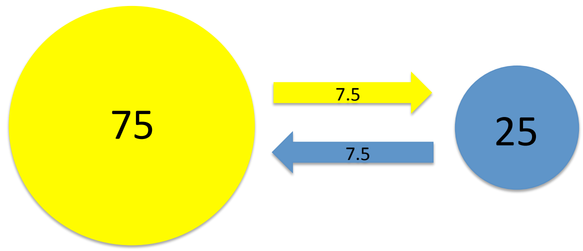


Figure 5.3: Equilibrium of a Markov Process

## The Ergodic Theorem

Our toy model of happy and sad people has four important features. First, each person had to be in one of a finite number of states. Second, the probability of moving from one state to another depended only on a person's current state and remained fixed throughout. And third, it was possible to move from any state to any other. Fourth, the process didn't produce a cycle in which every person alternated back and forth from happy to sad. The toy model exhibited an interesting property. It converged to a unique statistical equilibrium regardless of where the process was begun.

Mathematicians have shown a logical connection between these four assumptions and the convergence to a unique equilibrium. This connection can be made formal in what I will call the *Ergodic Theorem*.

## The Ergodic Theorem

A Markov Process has the following attributes

1. The state of the process at any point in time belongs to a finite set of possible states.
2. The state of the process at the next point in time depends only on its current state and does so according to fixed transition probabilities.
3. It is possible through a series of transitions to get from any one state to any other.
4. The system does not produce a deterministic cycle through a sequence of states.

**The Ergodic Theorem** *Any Markov Process converges to a unique statistical equilibrium that does not depend on the initial state of the process or any one time changes to the state during the history of the process.*

The Ergodic Theorem might also be called the "history doesn't matter" theorem. If its assumption are satisfied, then what has happened in the past doesn't influence the long run equilibrium. Yet, most of us believe that history does matter. So what gives? Let's look at the assumptions. Finite number of states? Seems pretty innocuous. No simple cycles? That makes sense too. Can get from any state to any other? Well, that could be problematic in some cases. Israeli and Palestinian people may not be able to get to the "we're best friends" state in light of events over the past few thousand years. So, there's one opening.

Fixed transition probabilities? Aha! Probably not. Here's where Markov processes don't fit history. Transition probabilities do change. When someone gets involved in a mutually supportive, constructive relationship their transitions from sad to happy become more probable. This observation doesn't mean that Markov processes are irrelevant to the study of historical processes. To the contrary, they tell us that events matter not if they change the state, but if they change the transition probabilities once we're in that state.

## Democratic Redux

Now that I've covered the basics of Markov theory, I want to use the model applied to the Freedom House data to show that every country probably won't become democratic. To do so, I have to parse that data a little more finely. In what follows, I approximate the probabilities that countries move between categories in any five year period. For the purposes of making the larger point, these crude approximations simplify the numerical calculations. If I were writing a research article, I would nail down the transition probabilities as closely as possible. I'd also test to see if they appeared to be fixed over the time period. Those caveats aside, let's assume the transition probabilities shown in the table below:

| | | **State in Next Period** | | |
|---|---|---|---|---|
| | | Free | Partly | Not Free |
| | Free | 95% | 5% | 0% |
| **Current State** | Partly | 10% | 80% | 10% |
| | Not Free | 5% | 15% | 80% |

If I seed the model with the data from 1975: twenty-seven percent free states, thirty-two percent partly free states, and forty-one percent not free states, then (assuming five year time steps) in 2010, I have forty-eight percent free states, thirty-one percent partly free states, and twenty-one percent not free states. The actual percentages are forty-six, thirty, and twenty-four respectively.

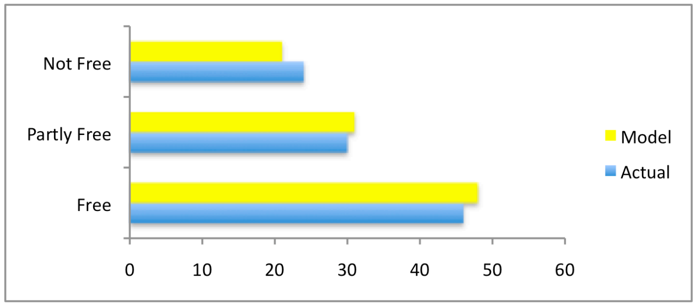


Figure 5.4: Markov Model vs Actual

Here's where the power of the Markov model enters. If we continue to apply these fixed transition probabilities, what happens? Well, in about fifty years, we should see only fifty-eight percent free countries, twenty-seven percent partly free, and fifteen percent not free. And in the long long run, we should expect sixty-two and a half percent free, twenty-five percent partly free, and twelve and a half percent not free.

I am not staking a claim that these predictions will be correct. I am merely stating that if countries continue to move between categories as they have been, we might more accurately expect about two-thirds of countries to be democratic then for all countries to be democratic. That doesn't mean that transition probabilities won't change. In fact, if we want a larger percentage of democracies, then we must figure out a way to change those probabilities.

## Fertility of the Markov Model

Lave and March emphasize that good models must be fertile – they must apply to multiple contexts. The Markov model proves especially fertile. It can be been applied in almost any setting in which a system moves probabilistically between states. Consider a person's health. We can categorize a person's health as being in some finite number of states. And, we can probably also assume that a person moves probabilistically between those states. At least in the short run, we can assume those probabilities remain nearly fixed, and thus, we can use the model to predict equilibrium health distributions.

Now, consider interventions such as drug protocols, behavioral changes, and surgeries. Typically, these change those transition probabilities. Evaluating the economic benefit of an intervention can be simplified by constructing a Markov model of a patient's health status both with and without the intervention.[3] If the intervention produces a better equilibrium, it's worth pursuing. If not, then the intervention should be abandoned. Of course, the challenge in constructing a useful model in such settings lies in how one defines the states.

A Markov model can also capture stock prices, where the price of a stock equals the state. Notice that here the implicit assumption is that the price in the next period does not depend on history. It only depends on the current price. In point of fact, stock prices may have some historisis – past prices may matter. If so, the Markov model can still be useful. By comparing actual price streams to those produced by the model, we can see how much history really matters.

Markov models can be used to recognize patterns in international crises (Schrodt 1998). The transitions that lead to war might look very different from the transitions that produce peace and reconciliation. To perform this sort of analysis, two different Markov models must be fit. The first uses data from cases where crises led to war. The second uses data in

[3] See Briggs and Sculper (1998) for an overview.

which crises were settled. If the two Markov models differ significantly, then we can look at a current pattern: bombing, hostage taking, no exchange of prisoners, escalated postering . . . , and ask which process was more likely to produce it: the process that leads to war or the process that doesn't.

## Outside the Box

Once we see that Markov models can be used to discriminate patterns a whole world of outside the box applications opens up. I'll describe just one: how Markov models can be used to help settle competing claims of authorship. That may seem an outrageous claim, but it's true. To get to that point, I first show how Markov Models can construct words, then get to sentences and books. Claude Shannon, a developer of information theory, constructed a Markov model that constructs words. In simplest form, the letters plus a "space" comprise the twenty-seven states of the model. The transition probabilities tell which letters are likely to follow another letter. For example, the letter $i$ would transition with high probability to the letter $e$ and to the letter $o$. It would also have a high transition to "space" so that it could create the word "I.".

True, this simple Markov model wouldn't be a particularly good writer. It would be more likely to spit out a nonsense sequence than a word "I athe pilly sared" than something deep like "whoso list to hunt," the opening words of Thomas Wyatt's famous sonnet to Anne Boelyn. Even so, this basic model reveals structure. It would encode the "i before e" rule by having the transition from $i$ to $e$ exceed the transition from $e$ to $i$. And, if we make the model even richer, by letting the state denote the previous two letters, then the rule that "i before e except after c" could be captured by having $ce$ transition to $i$ with high probability, but $ci$ not likely to transition to $e$. This two letter state model would have over seven hundred states and would be even better at constructing words.

Let's take this idea of having a Markov model write texts up one level of abstraction. Instead of having letters as states, let's have words as states. So, after the word "once" we might expect the words *in, upon,* or *twice*, but we would not expect the word *frequently.* To construct this model, we'd need a lot of computer memory (there are after all, a lot of words). And, we'd need some way of determining the probabilities that one word follows another. Doing this by hand would take a long time. Fortunately, we could automate it. We could write a program that reads books and keeps track of what words likely follow others. In fact, this has been done.

But which book? Different authors produce different patterns of word sequences. Thus, we could construct separate Markov model for Melville, Morrison, and Mao. This has been done as well.[4] These models wouldn't be sufficiently sophisticated to produce new texts by Melville or Dickens, but they could predict whether or not a disputed text was likely to have been written by a given author. We've now stumbled upon an unexpected use for Markov models: determining authorship. Take the case of the Federalist Papers, eighty-five essays written in 1787 and 1788 to convince New Yorkers to support the constitution. The authorship of twelve of these essays has long been a subject of dispute. Some say Hamilton wrote them and some attribute the works to Madison. The Markov model assigns them all to Madison.

We cannot say for sure that the model predicts correctly, but it makes giving Hamilton credit a more difficult proposition. I want to be clear. Models aren't always correct. They're very often wrong. But, what they do accomplish is focusing our thinking on some aspects of a situation and thinking through the logic clearly. Here that logic points to Madison.

[4] Khmelev and Tweedie (2001).

## Takeaways

We can now put the Markov model in our pocket. We know its assumptions: finite states, fixed transition probabilities, get to any state from any other, and no a simple cycle. We also know that given those assumptions that the process converges to a unique statistical equilibrium. That means that the process may keep churning but the statistical average remains unchanged. What do we do with this? Here are just a few takeaways.

**Takeaway #1** If we want to say that history matters, then we first have to say why history cannot be captured by a Markov model.

**Takeaway #2** In a system with many entities moving in an out of states, beware of inferring too much from initial linear trends. Pay attention to the long run equilibrium. We won't all be democracies.

**Takeaway #3** If a system is Markov, changing the current state has no long term impact. Making someone laugh won't change their long term demeanor, though it will give a temporary burst of joy.

**Takeaway #4** Fundamental change requires changing transition probabilities. It requires changing processes.

**Takeaway #5** Different processes produce distinct Markov model approximations. Therefore, we can fit Markov models to identify processes and to see the effects of interventions on long run equilibria.

To put a bow on this, California doesn't have one hundred million residents because while many many people did in fact move in, many others moved (including me!)