Conversation 8: Introduction to Markov Chains weather.com light

Winfried Just
Department of Mathematics, Ohio University

MATH3200: Applied Linear Algebra

Recap: What we learned in Conversation 7

- Real-world situations that involve not entirely predictable change can be modeled as *stochastic processes.*
- We will consider applications of linear algebra to *Markov chains,* which are one kind of stochastic processes.
- When constructing a Markov chain model of a real-world situation, we need to start by partitioning the possible observations of interest into non-overlapping *states* and specifying the meaning of each state.
- We will then be able to predict the probability of observing a given state at a given time in the future.
- *Probabilities* are numbers between 0 and 1. The probability of an event can be thought of as the proportion of times this event would occur if we conduct a large number of observations of very similar situations.

Recap of the beginning of the story

Bob got caught in a thunderstorm and our protagonists were wondering how one could use mathematics to predict the weather.

They decided to partition all possible weather patterns on a given day into two states:

1. State 1: sunny day.
2. State 2: rainy day.

Each state would then have a *probability*. For example, if we observe rain on 125 days of 2021 in a given location, the probability of a rainy day, or of state 2 rain in that location could then be estimated as $\frac{125}{365}$.

The vector $\vec{x} = [x_1, x_2] = \left[\frac{240}{365}, \frac{125}{365}\right]$ would be the *probability distribution* for the states on a randomly chosen day.

In a probability distribution, the individual probabilities x_i of the states must add up to 1.

Recap continued: Time steps and states

To build a *Markov chain,* we assume that time advances in discrete *steps:* $t = 0, 1, 2, \ldots$. One time step could last, for example, one second, one minute, or one hour. The meaning of one time step needs to be specified whenever we construct a Markov chain.

In our example, one time step would last one day.

Our protagonists established that in their numerical example, the probability of state 1, "sunny day," on a randomly chosen day, must be $\frac{240}{365}$.

Back to our story:

Denny: So now we can see that the probability of "sunny day tomorrow" must be $\frac{240}{365}$.

Question C7.3: Is Denny right?

This question leads us right into the essence of Markov chains. In this conversation, we will work out the answer.

Is today special?

Frank: I still don't buy it. Why could we assume that today is a "randomly chosen day"?

Denny: What's so special about today?
It's not Friday the 13^{th} or anything.

Theo: Frank means weatherwise. Let's assume for the sake of argument that we already made our observation today at noon.

Bob: I already did. And not for the sake of argument. It was awful, horrible! Very special indeed.

Cindy: I feel so sorry for you, Bob!

Theo: Now we know that today we are in state 2, rainy day.
And we are asking about the probability that tomorrow we will be in state 1, sunny day.

Denny: But we didn't send Bob out yet to make the observation for tomorrow. So, tomorrow would be like a "randomly chosen day" as it lies in the future.

Sunshine tomorrow? Take your bets.

Bob: Give me a break, Denny! How about if we send you out tomorrow without an umbrella? Rainy days tend to come in stretches, so good luck with your probability of $\frac{240}{365}$!

Question C8.1: What do you think about the probability of state 1, sunny day, for tomorrow given what happened to Bob today?

Alice: Our question is not about the overall probability of state 1, but about the probability that our Markov chain will *transition* from state 2 today (at time step t) into state 1 tomorrow (at time step $t+1$). If rainy days and sunny days come in stretches, this would be smaller than the probability that a sunny day at time t is followed by a sunny day at time $t+1$. It would also be smaller than the overall probability of a sunny day.

Cindy: So we will give you an umbrella before we send you out for the observation, Denny!

Transition probabilities

Theo: When constructing a Markov chain, we let p_{ij} denote the *transition probability* from state i to state j. This means p_{ij} is the probability that **if** the process is in state i at time t, **then** it will be in state j at time $t + 1$.

Alice: In our case, we were asking about the probability p_{21} of transitioning from state 2, rainy day, into state 1, sunny day.

Bob: What does all this have to do with linear algebra?

Theo: These probabilities can be organized into the *transition probability matrix* $\mathbf{P} = [p_{ij}]$ that has order $n \times n$, where n is the number of states. Together with the definitions of the time steps and the states, this matrix specifies the Markov chain.

Alice: In our example the number of states is $n = 2$, and the transition probability matrix takes the form $\mathbf{P} = \begin{bmatrix} p_{11} & p_{12} \\ p_{21} & p_{22} \end{bmatrix}$

The diagonal elements of $\mathbf{P}$

Denny: Wait! p_{11} would then be the probability that **if** today is a sunny day, state 1, **then** tomorrow will also be a sunny day, state 1, right?

Alice: Right!

Denny: So, in this case the weather stays in the same state and doesn't transition into anything!

Why would this be called a "transition probability"?

Frank: Glad you asked this question, Denny!

Theo: Any diagonal element p_{ii} does indeed give the probability that the process *remains* in state i at time $t+1$ if it is in state i at time t. You can think about it as transitioning from state i into the same state i.

Alice: Would you want us to say "transition and remaining probabilities," Denny?

Denny: Oh, no! Let's keep it short and simple. But why do you keep saying "time t" instead of "today" and "time $t+1$" instead of "tomorrow"?

t vs. $t+1$

Theo: This gives us greater flexibility. We could have:

- time t is today, time $t+1$ is tomorrow,
- time t is tomorrow, time $t+1$ is the day after tomorrow,
- time t is Friday, the 13^{th}, time $t+1$ is Saturday, the 14^{th},
- ... and so on.

The transition probabilities will always be the same under each interpretation.

Frank: This doesn't make sense!

Alice: Why not?

Frank: First of all, these probabilities may may fluctuate from season to season.

Second, it may matter also what happened earlier, whether today's rain was a one-off in the middle of a sunny period or one of a long stretch of rainy days.

The Markov property

Theo: Markov chains are mathematical models. Here we *assume* that the transition probabilities do not change over time and depend only on the current state, the state at time t, but not on prior history, that is on the states at times smaller than t. The latter assumption is called the *Markov property.*

Frank: But this is unrealistic. The weather doesn't have this Markov property.

Alice: You have a point here, Frank. The Markov property is a *simplifying assumption* that we make in our model so that the model becomes *mathematically tractable.*

Denny: Do you mean: "So that we still can calculate stuff with reasonable effort"? I'm all for that.

Alice: Yes. All models involve simplifying assumptions. We already made two of them when we decided to count time in steps of one day each and to partition all possible weather patterns into just two states. We do this with the hope that the Markov property will give us a *close enough approximation* to the actual dynamics of the weather so that our calculations will make reasonably accurate predictions.

An example of a transition probability matrix

Cindy: This all sounds very abstract. Can we look at some concrete examples?

Alice: Consider $\mathbf{P} = \begin{bmatrix} p_{11} & p_{12} \\ p_{21} & p_{22} \end{bmatrix} = \begin{bmatrix} 0.4 & 0.6 \\ 0.3 & 0.7 \end{bmatrix}$

Cindy: Can I think about these numbers in terms of tallying up observations?

Alice: Great idea! So suppose you pick 100 randomly chosen days in state 1, sunny days, and on each of these days you go out at noon the **next** day and make the observation. How many of those next days would you expect to be sunny, according to this matrix **P**?

Cindy: About 40, since $\frac{40}{100} = 0.4 = p_{11}$. And about 60 rainy days, as $\frac{60}{100} = 0.6 = p_{12}$.

Bob: And if I observe the weather on the next days after 100 rainy days, I should expect to observe about $100p_{21} = 30$ sunny days and $100p_{22} = 70$ rainy days. Or 30% sunny days and 70% rainy days, one might say. Now I can see what they mean!

Frank: Who are "they"?

Bob: Those percentages on weather.com! "15% chance of precipitation tomorrow" means that "about 15 out of 100 days where the weather is like today will be followed by a day with precipitation." And that means rain at this time of the year.

Cindy: You see Bob? $\frac{15}{100}$ is about 1 in 7, so when they give you this warning a whole week long, you should expect that it will rain on one day. You should take an umbrella. I don't want you to get all drenched and then sick!

Frank: But what would "weather is like today" mean?
No two days are exactly alike.

Alice: It could mean "in the same state of a certain Markov chain." Since weather.com has much more data and computational resources than we do, they could build a much larger Markov chain that takes into account also temperature, humidity, and atmospheric pressure.

Stochastic matrices

Denny: Let me see whether I got this "larger Markov chain" thing right. You mean "with more states"?

Alice: Yes, with many more states.

Denny: Let's not overdo the "many". How about 3 states? Could the transition probability matrix then be, for example

$$\mathbf{P} = \begin{bmatrix} p_{11} & p_{12} & p_{13} \\ p_{21} & p_{22} & p_{23} \\ p_{31} & p_{32} & p_{33} \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0.2 & 0.6 \\ 0.5 & 0.5 & 0.5 \end{bmatrix} \ ?$$

Alice: You got the order of the matrix right, but the numbers in your example don't work.

Theo: The transition probability matrix must be a *stochastic matrix,* that is a matrix where each element is a probability—

Denny: Which is the case in my example!

Theo: —and each row must sum up to 1.
That's not the case for rows 2 and 3 in your example.

Stochastic matrices, continued

Denny: Why do the rows need to sum up to 1?

Question C8.2: Why indeed? What would you tell Denny here?

Alice: Because when you are in state i, then you have to transition to **some** state j next. The probabilities p_{ij} for all states j are in row i, and if their sum is less than 1, it means there would need to be another possibility that is not represented by a state in your model.

Theo: And if the probabilities sum up to more than 1, you would need to sometimes go to several states at once, which violates our assumption that there is no overlap between different states.

Denny: I see. Would the following example work?

$$\mathbf{P} = \begin{bmatrix} p_{11} & p_{12} & p_{13} \\ p_{21} & p_{22} & p_{23} \\ p_{31} & p_{32} & p_{33} \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0.1 & 0.2 & 0.7 \\ 0 & 0.5 & 0.5 \end{bmatrix}$$

Alice: Yes, this would be a good example.

Take-home message: Markov chains

- Real-world situations that involve not entirely predictable change can be modeled as *stochastic processes.*
- Here we introduced *Markov chains,* which are one kind of stochastic processes.
- In Markov chains, time is assumed to proceed in discrete *steps,* and the real-world situations of interest are categorized into nonoverlapping *states.*
- When constructing a Markov chain model of a real-world situation, we need to start by specifying the meaning of each state and the meaning of one time step.

Take-home message: Transition probabilities

- In Markov chains, we are interested in the probabilities p_{ij} that when the system is in state i at time t, it will be in state j at time $t + 1$.
- These probabilities p_{ij} are called the *transition probabilities.* In Markov chains, they do not depend on t, and also do not depend on how prior history, that is, how the system reached state i.
- The *transition probability matrix* $\mathbf{P} = [p_{ij}]$ has order $n \times n$, where n is the number of states, and is a *stochastic matrix,* which means that each of its rows sums up to 1.

- In Markov chains, we are interested in the probabilities p_{ij} that when the system is in state i at time t, it will be in state j at time $t + 1$.
- These probabilities p_{ij} are called the *transition probabilities.* In Markov chains, they do not depend on t, and also do not depend on how prior history, that is, how the system reached state i.