Conversation 7: Introduction to Probabilities

Winfried Just
Department of Mathematics, Ohio University

MATH3200: Applied Linear Algebra

Bob checked weather.com and . . .

Alice: Hi Bob! What happened to you?

Bob: There was a thunderstorm and I got all drenched.

Cindy: Hope you will not catch a cold.
You should have taken an umbrella!

Bob: I checked weather.com before I left and it showed sunshine.

Frank: Sue them.

Denny: Don't trust them.
They just give you some fake info to make you look at their ads.

Alice: What, exactly, did their website show?

Bob: Precipitation 15%. Means it won't rain.

Cindy: But it means there is a small chance it might, so you should have taken the umbrella, Bob. Now I am all worried that you might catch a cold.

Frank: Don't be so naive, Cindy! They tell you it won't rain and then they put these percentages to cover their— (Alice frowns)
—back so that you can't sue them.

Theo: No, Cindy is right. "Chance of precipitation 15%" is a mathematically precise way of saying that there is a small chance that it might rain.

Denny: Come on, Theo! You cannot use math to predict the weather. It keeps changing and you cannot be sure how.

Theo: Yes you can. You can model it as a *stochastic process.*

Denny: Scholastic progress? What's that? I thought scholastic was something medieval and not progressive at all.

Theo: "Process." Means we are studying a situation that is changing. And "stochastic," which means that there is some unpredictabaility of what will happen next, exactly as you told us about the weather.

Denny: So if it's unpredictable, how can you use math?

Theo: Stochastic process models allow us to study how the probabilities of certain events change over time.

Denny: This all sounds very scholastic to me, if you excuse my Latin. Can you explain it in plain English?

Theo: An *event* is simply something that might or might not happen.

Cindy: You mean, like rain today or sunshine tomorrow, right?

Theo: Excellent examples, Cindy! Now the *probability of an event* is a number between zero and one that signifies the frequency with which that event occurs. We can think about determining it by observing many similar situations, keeping a checklist of in how many of these situations the event occurred, and dividing by the total number of observations.

Cindy: So, like, if we observe rain on 125 days of 2021 in a given location, the probability of rain in that location would be $\frac{125}{365}$?

Theo: Yes. To be precise, this would give us an estimate of the probability of rain on a randomly chosen day in that location.

Events and probabilities

Bob: Wait! But I got already drenched today, that's for sure, and there is no $\frac{125}{365}$ about that!

Theo: You are talking about an event that has probability 1, which signifies certainty about its occurence.

Bob: That's for sure, but how does it fit with your checklist?

Theo: You are talking about an event that occurred in the past and that you have already observed. We would use the checklist to estimate probabilities of events that we have not yet observed, especially events that might or might not occur in the future.

Cindy: Like the probability of sunshine tomorrow!

Denny: So if there are 125 days with rain, then there are $365 - 125 = 240$ days with sunshine, and the probability of sunshine tomorrow would be $\frac{240}{365}$.

Question C7.1: What do you think about Denny's reasoning?

We will see.

What's linear algebra got to do with it?

Frank: I don't buy this! What about days where it rains in the morning and the sun shines in the afternoon?

Denny: Good point! So there may be more than $365 - 125 = 240$ days with sunshine. We could then say that the probability of sunshine tomorrow would be at least $\frac{240}{365}$, but we couldn't calculate its value.

Frank: I still don't buy it. What about dry, but overcast days?

Denny: Then ... then we can't say anything. That was my point: When there is uncertainty involved, math goes out the window.

Theo: There is a branch of mathematics, called *probability theory*, that deals with quantifying uncertainty.

Bob: Come on Theo! This is a course on linear algebra, and probability theory is not a prerequisite. Let's focus on matrix multiplication instead.

Alice: But there are stochastic processes, called *Markov chains,* that are all based on matrix multiplication. Can you tell us about them, Theo?

Markov chains: Time steps and states

Theo: In 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.

Cindy: When we want to know about the weather tomorrow, one time step would last one day, right?

Theo: Right! Now we need to partition all possible weather patterns during a day into a finite number of *states* that also need to be specified. For simplicity, let us take as our states here:

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

Frank: So how about a day when it rains in the morning and the sun shines in the afternoon?

Alice: Good question Frank!
We need to define the states so that there is no overlap.

Denny: Simple. We go out each day at noon and look. Or make an observation, as Theo would put it. If it rains, state 2, rainy day. If not, state 1, sunny day!

Bob: But how about overcast and dry days?
Should we have a third state for them?

Theo: We might, but let's keep this as simple as possible and restrict ourselves to two states here.

Denny: Yeah. Let's keep it simple.

Frank: Shouldn't we call state 1 "not rainy" then?

Theo: That would be mathematically more precise. But the name is just a label we attach to the observation, so it can be short and doesn't need to tell the whole story, as long as we have clearly specified the precise definition. Denny has done this for us.

Cindy: And "sunny" sounds so much more cheerful!

Denny's calculations revisited

Cindy: So, if I went out in 2021 each day at noon, found that it rained on 125 of them, then my checklist would give me a probability of $\frac{125}{365}$ for state 2, rainy day, right?

Theo: Right!

Denny: And there would now be $365 - 125 = 240$ days that we decided to call "sunny." So this gives a probability of $\frac{240}{365}$, or $1 - \frac{125}{365}$ for state 1, sunny day! As I told you.

Question C7.2: Is Denny right about the probability of state 1?

Theo: Exactly! 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. Denny has used this fact in his calculation.

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

Question C7.3: Is Denny right?

Alice: Let's discuss this question next time.

Take-home message

- 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.