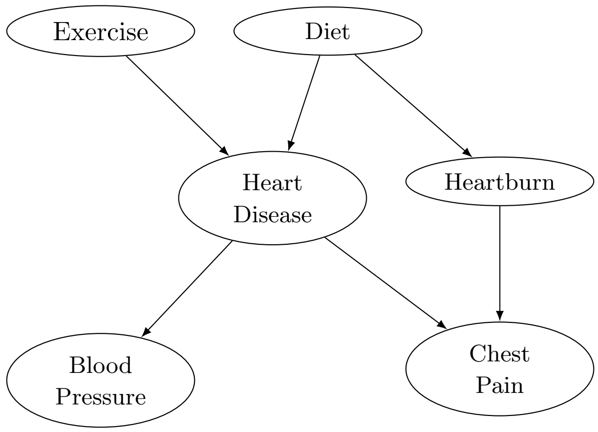
Question 1. Review: give several conditional independences and conditional dependencies in the graph below.



Question 2. Suppose you are using Bayesian Networks to write software for a medical center. Which of these random variables can the medical doctors at your center directly, or very closely, observe? Which are the most difficult for them to observe?

Reasoning Over Time

Here is a problem you might want to solve. Suppose the weather every day is exactly one of "rainy" or "sunny" and that tomorrow's weather depends only on today's. Today is rainy[1]. For any given day, if it is rainy, there is a 50% chance that tomorrow is rainy. If any given day is sunny, tomorrow is sunny 80% of the time.

If today is rainy, what is the probability that it is rainy ten days from now? Note that the days in between may or may not be sunny.

Question 3. How would you solve this in probability class?

[1] this is not necessarily a realistic problem in Southern California.

We call this sort of problem a **Stationary Markov Chain** modeling. In this, we are given ordered random variables $X_1, X_2, \cdots, X_t, \cdots, X_T$, called *states*, along with **Transition probabilities** for describing how the state at time $t-1$ changes to the state at time t,

$$P(X_t = \mathsf{value}' | X_{t-1} = \mathsf{value})$$

and also **initial probabilities** for describing the initial state at time $t = 1$.

$$P(X_1 = \mathsf{value})$$

We sometimes abbreviate transition probabilities as a_{ij} and initial probabilities as π_i

In order to work with this model, we need to compute

$$P(X_1 = x_1, X_2 = x_2, \cdots, X_T = x_T)$$

We can use the Markov property to factor

$$P(X_1 = x_1, X_2 = x_2, \cdots, X_T = x_T) = P(X_1 = x_1) \prod_{t=2}^{T} P(X_t = x_t | X_{t-1} = x_{t-1})$$

We can then use a technique called "Maximum likelihood estimation" that is *way* beyond the scope of this course to get this:

$$\pi_i = \frac{\#\text{of sequences starting with } i}{\#\text{of sequences}}$$

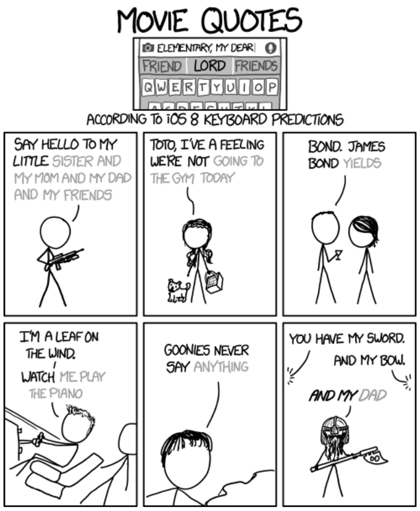
and

$$a_{ij} = \frac{\#\text{of transitions starting with } i \text{ but ending with } j}{\#\text{of transitions starting with } i}$$

Application: Text Prediction

In this application, "states" are words. The probability of the "next word" depends on the previous word.

Question 4. Feel free to assume all sentences are some fixed length T and that we have a file with **MANY** sentences. How do we estimate the probabilities? You do not need to write code (in R or otherwise); instead, imagine you were giving instructions to a human to hand-compute these probabilities from that document.



XKCD # 1427; More actual results:

- ‘Hello. My name is Inigo Montoya. You [are the best. The best thing ever]’,
- ‘Revenge is a dish best served [by a group of people in my room]’, and
- ‘They may take our lives, but they’ll never take our [money].’

Question 5. The above comic shows [humorous] incorrect text prediction for nine famous movie quotes. How many are the correct famous quote? Why do you suppose some are incorrect?

Question 6. How would you make text prediction a more complicated program, while also giving better predictions? What sorts of changes to the previous description we gave would you want to make?

Observations and Causes

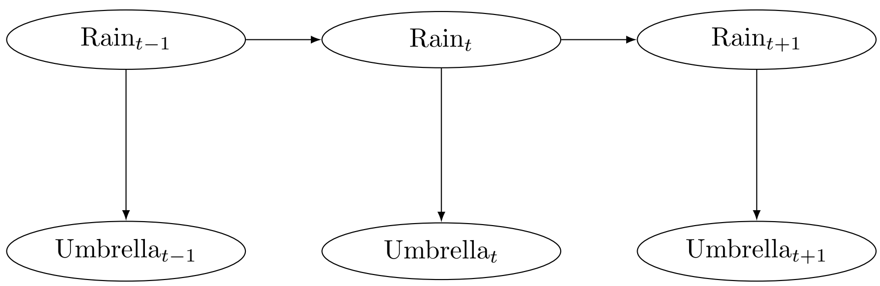
Here is a more common type of problem dealing with Markov Chains. There is an underlying event that we cannot observe directly. This is similar to heart disease in the opening example: we can see symptoms, but we don’t necessarily know the ground truth of it.

The following is a very simplified problem to illustrate the ideas behind this type of machine learning. Suppose you live in a building and never go outside, nor do you observe the outside. Perhaps the building is underground and you watch the elevator. There is one resident who goes outside daily, and you believe this person knows something about the weather. You cannot ask about the weather, but you notice some days this person carries an umbrella upon leaving and other days not. You aren't present on return to check if the umbrella is wet or not. You want to figure out which day(s) it rained outside. For this example, suppose an umbrella is only used for rain, not for sun.

The weather is a Markov process, so the probability it rains tomorrow depends on today's rain or not. However, whether or not our friend takes an umbrella outside on any given day depends on *that day's weather* – regardless of why it is or isn't rainy, and conditionally independent (given the weather) of whether or not an umbrella was taken yesterday.

The following network illustrates a segment of days:

R_{t-1}	$P(R_t \mid R_{t-1})$
T	0.5
F	0.2



R_t	$P(U_t \mid R_t)$
T	0.9
F	0.2

Question 7. The probabilities in each table do not add up to 1 (= 100%). Is this an error? Why or why not?

Question 8. Suppose I observe that my neighbor had the umbrella on days 1, 2, 4, and 5 last week. What is the most likely sequence of weather? *Note that it isn't necessarily rain on those days, sunny on day 3.*

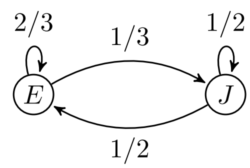
For purposes of this problem, we believe it was 50/50 for Rainy/Sunny on "day 0" (the day before the first day we care about).

Extra Practice

E	$p(E \mid X = \text{Erlang})$
happy	4/5
angry	1/5

E	$p(E \mid X = \text{JavaScript})$
happy	1/4
angry	3/4

In the diagram below, E denotes Erlang and J denotes JavaScript



You notice that I am happy for days one and two and angry for day three. What is the most likely sequence of programming languages I used? *Note that it is not necessarily Erlang the first two days and JavaScript on day three.*