COMP 150PP Class Exercise: Playing With Dice

September 7, 2016

This exercise will introduce us to *probabilistic inference*. We may also get into a little bit of machine-learning terminology.

At the start of class, please close your eyes and *choose a die from the bowl* of dice. *Don't let anyone see which die you've chosen;* just stick it in your pocket.

A die is identified by its number of sides. The bowl includes these sorts of dice: d6, d8, d12, and d20. (The numbers on an N-sided die range from 1 to N, *except* for the d10, which may be numbered from 0 to 9 or from 00 to 90. I've left the d10's and the d4's out of this experiment.)

The experiment

Find a partner so that the two of you have a *pair* of dice. Now run the following probabilistic experiment:

1. Pull a desk to the edge of the room facing outward, so nobody can see what you are working on.

2. Take a blank *tally sheet* and put both your names on it. The tally sheet should be divided into three columns, thus:

Total $<$ 16	Total $=$ 16	Total $>$ 16
...	...	...

3. Take your pair of dice and throw the pair 30 times. Each time, add the total of the numbers on the two dice and put a mark on your tally sheet:

   - If the total is less than 16, make a mark in the left column
   - If the total is exactly 16, make a mark in the middle column
   - If the total is more than 16, make a mark in the right column

   Please make your marks in the classic "prisoner's groups" of five (four verticals and a diagonal).

4. Once you have completed your tally sheet, please

    - Hide your dice (in your pocket?)
    - Write, at the bottom of your tally sheet, how many marks appear in each column
    - Give me your tally sheet
    - Return to the group

Here's the problem of the day: if I see only your tally sheet, can I tell what dice you have? This problem is an *inference* problem.

Inferring unknowns from knowns

Let us call the contents of the bowl B. The contents are known initially, and from the known contents, if we assume all dice are equally likely to be drawn, we can compute a probability distribution of dice D. The dice are drawn "without replacement" (as the statisticians say), and my computer code accounts for changes in the distribution as dice are drawn. But in our calculations in class, let us assume that dice are drawn "with replacement", so that D is the distribution for all the dice, and D does not change.

Here are the initial contents of the bowl:

Number of sides	Number of those dice in the bowl
d4	—
d6	9
d8	9
d10	—
d12	14
d20	14

I don't get to see what pair of dice you picked, so that's an unknown. And nobody except you gets to see the totals thrown on the dice, so those are unknowns too. But we do get to see your tally sheet, so the number of marks in each column is known.

Here's the inference problem: *given your tally sheet, what dice do you have?*

Discussion questions

1. Please sketch how would you write a program to simulate the experiment with the tally sheet. Make it a quick sketch, and imagine you get to use

the programming language of your choice.

2. What ideas so you have about how would you write a program to solve the tally-sheet inference problem?

3. Now I'll ask you to switch from computation to mathematical notation.

   The rest of this question is about *notation* for probabilities. *Without calculating* any numbers, please answer these questions:

   A. How would you notate the probability of drawing a d6 and a d12 and throwing a total of 11?

   B. Refactor your formula from part A to include the *conditional* probability of throwing a total of 11 *given* the assumption that you drew a d6 and a d12.

   C. Refactor your formula from part A to include the *conditional* probability that you drew a d6 and a d12, *given* the observation that you threw a total of 11.

   D. Briefly discuss how you might compute the probabilities and the conditional probabilities from parts A, B, and C.

      Hint: The easiest computation here is the one in part B. Most people find it easier to count events than to reason about probabilities. So given a pair of dice (say d6 and d12), you might imagine a set of possible worlds: one for each possible outcome of each throw of each die (say $6 \times 12 = 72$ worlds in total). Count the number of times the dice total 11, count the total number of possible worlds, and the ratio is your conditional probability. Try applying the same ideas to parts A and C.

   E. Using the elements of your previous formulas, please write down a formula for the conditional probability that you drew a d6 and a d12, given that the total of one throw was more than 16.

      Hint: at some point you will get your formula down to a bare probability that the total was more than 16. A key step is to rewrite that probability as a sum:

      $$P(total > 16) = P(total > 16 \wedge (d6, d12)) + P(total > 16 \wedge \neg(d6, d12))$$

      What's the game here? We want to split all outcomes of totals > 16 into two parts: those produced with a d6 and a d12, and those produced with all other pairs of dice. Introducing that split is the key to computing the *weight of evidence* that the total introduces in favor of the hypothesis that you drew a d6 and a d12.

Weight of evidence, prior belief, posterior belief

The *log odds* of a hypothesis H is the log of the ratio of the probabilities

$$10 \times \log_{10} \frac{P(H)}{P(\neg H)}$$

The unit "ten times the base-10 logarithm" is called the *deciban*; according to I. J. Good, it was invented by Alan Turing.

Log odds gives exactly the same information as probability. In fact, there is a (monotone) bijection between the two. So why use log odds? Because it gives us a more effective tool for distinguishing very large or very small probabilities.[1]

- An even-money bet has log odds of 0 decibans.
- An event that happens ten times out of eleven (10-to-1 odds in favor) has log odds of 10 decibans. A hundred to 1 is 20 decibans, and a million to 1 is 60 decibans.
- The difference between a 99% chance and a 99.5% chance may not seem like much, and it's not—only about 3 decibans, which is hardly enough for people to distinguish. But the difference between 99.5% and 99.5%, which in probability looks like it might be even smaller, is actually 7 decibans, which is significantly more convincing.

Here's a table from Harold Jeffreys (via Wikipedia) saying how people view absolute values of probabilities in decibans:

0 to 5	Barely worth mentioning
5 to 10	Substantial
10 to 15	Strong
15 to 20	Very strong
above 20	Decisive

And another table from Kass and Raftery:

0 to 5	Barely worth mentioning
5 to 13	Positive
13 to 22	Strong
above 22	Very strong

[1] There's another reason to take logarithms of probabilities: if we want to represent probabilities in the computer, we'll use bits more effectively if we store the logarithms. Multiplying such probabilities is easy (add the logarithms); adding them requires a trick.

Wait a minute? I thought we were talking about *weight of evidence*! What's that about? Weight of evidence quantifies how you should *modify your beliefs in response to observation.* You posterior believe should be your prior belief multiplied by the weight of evidence.

To illustrate weight of evidence, I'm going to pick the most extreme possible pair of dice: two d6's. And I'm going to consider a much simpler experiment in which you throw the dice just once.

Your chance of drawing two d6's from the bowl under four percent (odds about 30 to 1 against, log odds about 14 decibans against). When you throw your dice you get a total at most 12, so you put a mark in the left column. What should I believe now? *If* you have two d6's then the probability of your marking the left column is 100%. But if you have anything except two d6's, the probability of your marking the left column is about 65%.[2] Given that I've observed a mark in the left column, the weight of evidence in favor of the hypothesis that you have two d6's is 10 times the log of the ratio of the two probabilities, or about 2 decibans. With that weight of evidence, I'm going to change my mind and decide that the log odds of your having two d6's is only $-14+2=-12$ decibans, only about 16 to 1 against. Those 2 decibans make it $10^{\frac{2}{10}} = 1.5$ times more likely that you have the d6's.

Terminology:

- The log odds of -14 decibans that you have two d6's represent my *prior belief.* The sum total of my prior beliefs about all possible pairs of dice you might have constitute a *prior distribution.*
- Given the observation, I can add the weight of evidence in favor of the "two d6's" hypothesis to get a *posterior belief*: log odds of -12 that you have two d6's. I should modify my beliefs about all other possible pairs of dice as well; that would give me a *posterior distribution.*
- The weight of evidence is multiplicative:

  $$Posterior = WofE \times Prior$$

  . But normally we take logarithms and add.

  The formula for the weight of evidence of an observation O in favor of a hypothesis H is

  $$\frac{P(O|H)}{P(O|\neg H)}$$

The tally sheet has results from 30 throws, not just 1. Wouldn't it be great if we could solve it by adding up the weight of evidence for each mark, one

[2] The exact probability depends on what dice were left in the bowl when you drew. The figure of 65% assumes you drew first.

at a time? This idea actually works, but unfortunately it's not as simple as taking 2 decibans and multiplying by the number of marks in the left column.[3] Unfortunately this idea doesn't work. Because the marks are all made by the same dice, the probability of the second mark depends on the first mark:

- Probability of making the first mark in the left column: about 66%
- Probability of making the second mark in the left column: about 66%
- Probability of making the second mark in the left column, given that the first mark is also in the left column: about 74%

A single mark in the left column adds 1.9 decibans to the log odds of your having two d6's, but two marks in the left column add only 1.4 decibans—once we see that first mark, the second one doesn't carry as much evidentiary weight.

Ultimately, solving tally-sheet inference involves combinatorics. We'll have the chance to consider a number of different solutions, and in the process, we'll start to develop abstractions suitable for probabilistic programming languages.

Advanced questions

Here are some questions ranging from hard to insanely hard. The purpose of these questions is to help convince you that a domain-specific language for probabilistic reasoning could be mighty useful:

a) I'm going to look at your tally sheet and guess which dice you have. If I guess correctly, you have to pay me a dollar. Then we do it again, as many times as I like. What's the minimum amount you should charge me for the privilege of playing this game, so that in the long run, you expect not to lose money?[4]

b) Associating the middle tally column with a totals of 16 seems arbitrary. Why 16? What number should I pick to maximize my chances of guessing correctly?

c) Matthias Scheutz invents a robot that draws two dice from the bowl, throws the dice 30 times, produces a tally sheet, and puts the dice back. If the bowl contains 100 dice, how many tally sheets does the robot need to produce for us to be 90% confident that we can guess the number of d6's in the bowl?

[3] We'd need to account for the right column as well; if you throw the dice once and mark the middle or right columns, that's infinite evidence *against* the hypothesis that you have two d6's.

[4] We ignore the value of your time spent throwing dice and of my time spent guessing.