Chris Piech
CS 109

Section #5
Feb 13, 2025

Section 5: Probabilistic Models

With questions by Chris

1 Warmup

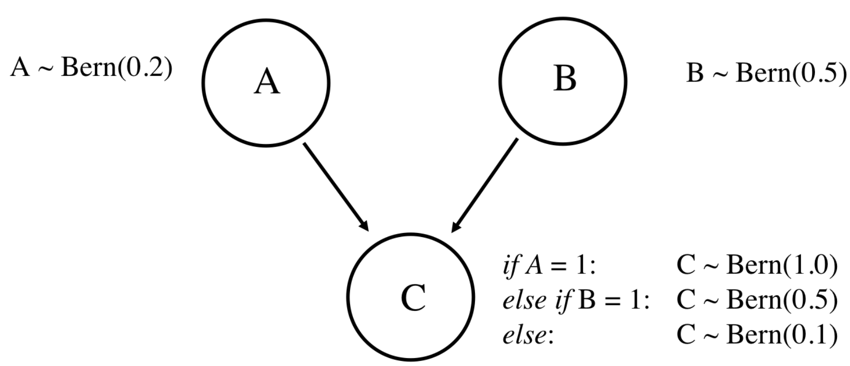
What is a probabilistic model with multiple random variables? What does the term inference mean? What do you call the probability of an assignment to all variables in a probabilistic model? Why is that useful? Why can it be hard to represent?

A probabilistic model is a way of defining how a set of many random variables vary together. A probabilistic model should describe the relationships between many random variables, including what RVs are independent from each other, and what the joint probabilities are for assignments to all the RVs at once. Inference is the act of computing an updated belief in one (or more) random variables, based on one or more observations. The probability of an assignment to all variables in a probabilistic model is called the joint. The joint is the "and" between all the RVs and can be used to solve any inference task. The number of ways of assigning values to variables is exponential in the number of random variables, so the size of the joint (if represented as a table) can be huge.

2 Understanding Bayes Nets

	A = 0		**A = 1**	
	B = 0	B = 1	B = 0	B = 1
C = 0	0.36	0.20	0.00	0.00
C = 1	0.04	0.20	0.10	0.10

The **joint probability table (above)** for random variables A, B and C is equivalent to the **bayesian network (below)**. Both give the probability of any combination of the random variables. In the Bayes network the probability of each random variable is provided given its causal parents.



a. Use the bayesian network to explain why $P(A = 0, B = 1, C = 1) = 0.20$.

> We can intentionally break down this joint probability using the Chain Rule, so that we are left with probabilities of events that the bayes net directly gives us: $P(A = 0, B = 1, C = 1) = P(A = 0)P(B = 1)P(C = 1|A = 0, B = 1) = 0.8 * 0.5 * 0.5 = 0.2$.

b. What is $P(A = 1|C = 1)$?

> Using the table, we see that
>
> $$P(A = 1|C = 1) = \frac{0.1 + 0.1}{0.1 + 0.1 + 0.2 + 0.04} = \frac{0.2}{0.44} = \frac{5}{11}$$
>
> If we wanted to find this probability using the bayes net instead, it would be possible, but it would require using Bayes' Theorem, since the structure of the network makes it much easier to think about $P(C = 1|A = 1)$ than $P(A = 1|C = 1)$.

c. Is A independent of B? Explain your answer.

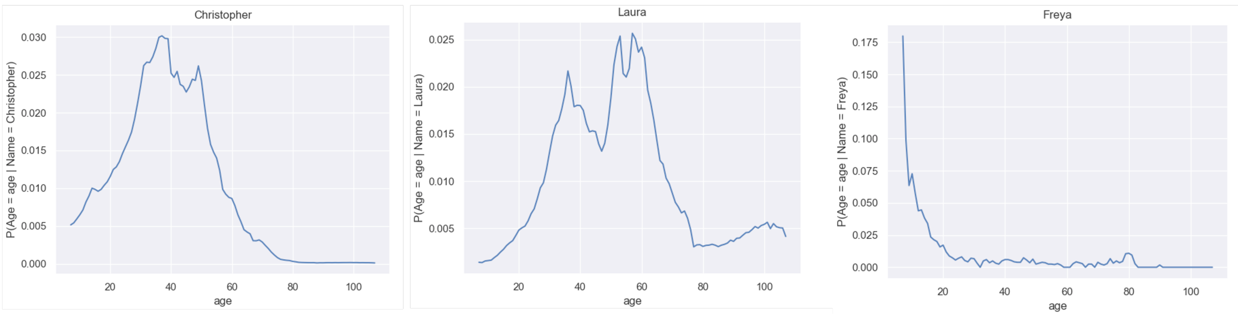
> Yes. This follows directly from the structure of the bayesian network, because A and B have no shared ancestors. Alternatively, note that from the table or the network, we can calculate that $P(A = a, B = b) = P(A = a)P(B = b)$, which satisfies the definition of independence.

d. Is A independent of B **given** $C = 1$? Explain your answer.

> No. From the table, we can see that $P(B = 1|A = 0, C = 1) = \frac{0.2}{0.04+0.2} \neq P(B = 1|A = 1, C = 1) = \frac{0.1}{0.1+0.1}$. So given $C = 1$, knowing the value of A informs us about the value of B, and therefore A and B are not conditionally independent given C.
>
> This probably feels counterintuitive! Here's an example of what these random variables could be to illustrate how this is possible. Let A, B, and C each be Bernoulli RVs as in the bayes net. C is 1 if you are holding a fresh mandarin, 0 otherwise. A is 1 only if you recently received a mandarin from Chris in lecture, and B is 1 only if you recently bought mandarins at the store. If you have a mandarin ($C = 1$), it's very likely that you got it either from Chris ($A = 1$) or from the store ($B = 1$). So then, if you have a mandarin and tell me you haven't been to the store recently ($B = 0$), that leaks information about A, because I can guess that you probably received your mandarin from Chris!
>
> This phenomenon is sometimes called 'Explaining Away' if you're curious to read more.

3 Name2Age Inference

What is the probability distribution of someone's age given just their name? Here are a few example for the names 'Christopher' 'Laura' and 'Freya':



The U.S. Government released a dataset of the frequencies, by year, of all given names recorded in U.S. births at least 5 times. You can access this data via the function `get_count`(`name`, `year`) which returns the number of babies named `name` born in `year`. Since this data provides the joint distribution, it can be used to solve inference problems. The code and data are available here: `http://web.stanford.edu/class/cs109/section/5/babynames.zip`

Write a function in pseudocode that 1) takes in a name and infers the conditional distribution $P(\text{Age} = \text{age}|\text{Name} = \text{name})$ across all of the ages covered by the dataset, and 2) plots this conditional probability function (see the plots above as examples).

```
def run_name_query(name, years_list):
```

See this chapter in the course reader: `https://probabilityforcs.firebaseapp.com/book/name2age`
Code solutions and more are in the Colab notebook linked on the course website.