Chris Piech
CS 109

Section #5
Nov 3, 2021

Section 5: Probabilistic Models

With questions by Chris

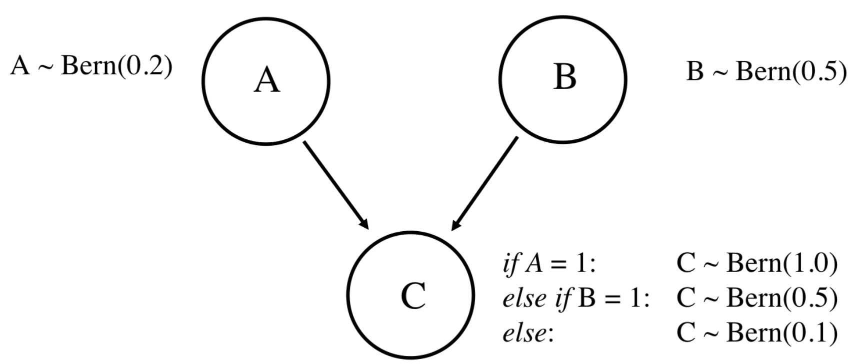
1. **Warmup**

   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 the relationship between many random variables. Inference is the act of computing a belief in one (or more) variables based on an observation. The probability of an assignment to all variables in a probabilistic model is called the joint. The joint 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.

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 $P(A = 0, B = 1, C = 1) = 0.20$

> $P(A = 0, B = 1, C = 1) = P(A = 0)P(B = 1)P(C = 1|A = 0, B = 1) = 0.8 * 0.5 * 0.5 = 0.2$.

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

Using the table, we see that

$$P(A = 1|C = 1) = \frac{0.1 + 0.1}{0.1 + 0.1 + 0.2 + 0.04} = \frac{0.2}{0.44} = \frac{5}{11}$$

(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 $P(A = a, B = b) = P(A = a)P(B = b)$, which satisfies the definition of independence.

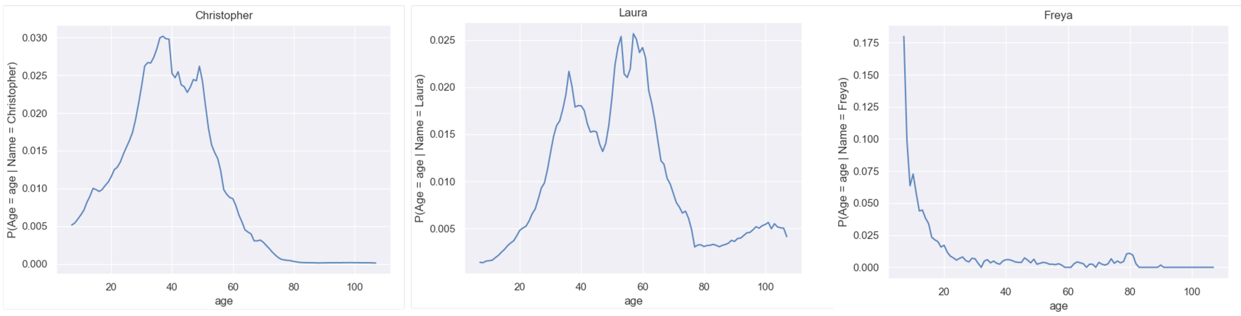
(d) Is A independent of B **given** $C = 1$? Explain your answer.

No. From the table, we can see that $P(B = 1|A = 0, C = 1) = \frac{0.2}{0.4+0.2} \neq P(B = 1|A = 1, C = 1) = \frac{0.1}{0.1+0.1}$. 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.

Note: 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, list_of_all_years):
```

`https://chrispiech.github.io/probabilityForComputerScientists/en/examples/name2age/`

4. **Beta Distribution**

An item on an online store has 10 ratings. 9 likes and 1 dislike. What is your belief that the true value of p is < 0.8? Assume a Uniform prior for your belief in the true probability and use `scipy.stats.beta.cdf(x, a, b)`

Let X be a random variable for the probability that you like the item. $X \in [0, 1]$.
$X \sim \text{Beta}(a = 10, b = 2)$ since $a = \text{successes} + 1$ and $b = \text{fails} + 1$ with a uniform prior

$$\begin{aligned} P(X < 0.8) &= F_X(0.8) \\ &= \texttt{scipy.stats.beta.cdf(0.8, 10, 2)} \\ &= 0.322 \end{aligned}$$