

Section 5: Probabilistic Models

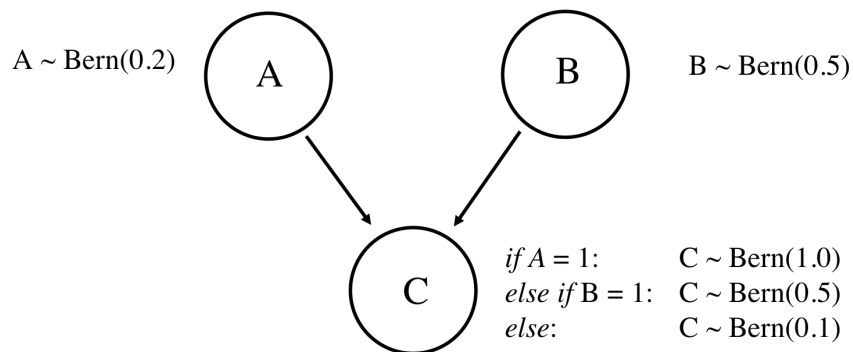
With questions by Chris

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?

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

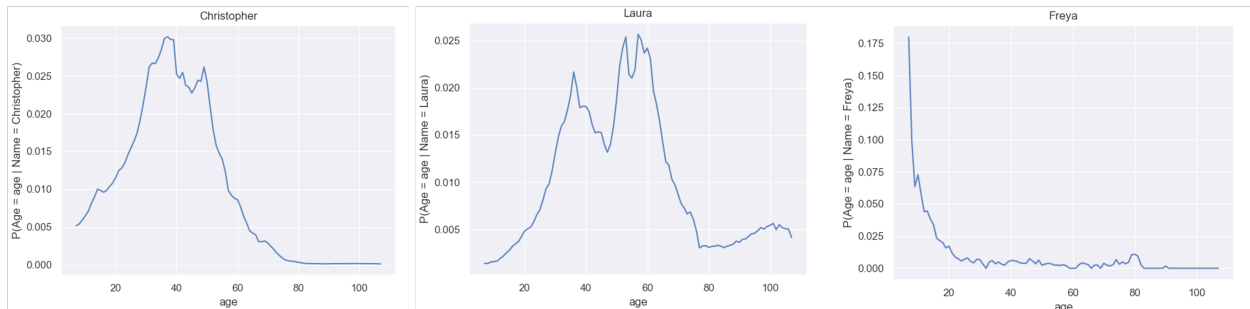


The **joint probability table (above)** for random variables A , B and C is equivalent to the **bayesian network (below)**. Both give the probability of any combination of the random variables. In the Bayes network the probability of each random variable is provided given its causal parents.

- (a) Use the bayesian network to explain why $P(A = 0, B = 1, C = 1) = 0.20$
- (b) What is $P(A = 1|C = 1)$?
- (c) Is A independent of B ? Explain your answer.
- (d) Is A independent of B **given** $C = 1$? Explain your answer.

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 US Government released a dataset on the relative frequency of given names in the population of U.S. births where the individual has a Social Security Number. To safeguard privacy, they restrict their list of names to those with at least 5 occurrences. You can access this data via a function `get_count(name, year)` which returns the number of babies born in a particular year with a particular name. Since this data provides the joint distribution, implicitly, 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>

Use this function to infer the conditional distribution $P(\text{Age} = \text{age} | \text{Name} = \text{name})$.

Based on your derivation write a function that could plot the conditional probability function:

```
def run_name_query(query , all_years , data ):
```

4. Beta Distribution

An item on an online store has 10 ratings. 9 likes and 1 dislike. Is the probability that we like the item truly $p \approx \frac{9}{10}$? There are not that many ratings and as a result we should have more uncertainty in our estimate of p than if we had, say, 100 ratings. 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)`