1 Chain Rule of Probability

1.1 Two Variable Case

For two random variables A and B, the probability distribution over them is.

$$P(A, B) = P(A|B)P(B)$$

Which is equivalent to

$$P(A, B) = P(B|A)P(A)$$

Using these two rules, we can derive Bayes Rule:

$$\begin{aligned} P(A, B) &= P(A, B) \\ P(A|B)P(B) &= P(B|A)P(A) \\ P(A|B) &= \frac{P(B|A)P(A)}{P(B)} \end{aligned}$$

1.2 Three Variable Case

What about for three random variables A, B, C?

$$P(A, B, C) = ?$$

Well assume X to be random variable which can take on all values that the A, B pair can take on, with equal probability distributions. Using the rule from two variable case, we can now write

$$\begin{aligned} P(A, B, C) &= P(X, C) \\ &= P(C|X)P(X) \end{aligned}$$

After replacing X with (A, B) again, we get

$$\begin{aligned} P(A, B, C) &= P(C|X)P(X) \\ &= P(C|A, B)P(A, B) \\ &= P(C|A, B)P(B|A)P(A) \end{aligned}$$

1.3 General Form

This discovery is called the chain rule of probability. It's general form states that:

$$P(A_n, A_{n-1}, ...A_2, A_1) = P(A_n|A_{n-1}, ..., A_1)P(A_{n-1}|A_{n-2}..., A_1)...P(A_2|A_1)P(A_1)$$

1.4 Relating this to Conditional Independence

Recall the idea of conditional independence. When A is conditionally independent of B given C, we can say that:

$$P(A|B, C) = P(A|C)$$

We can take advantage of conditional independence between variables to simplify our general form of chain rule. For example, again consider random variables A, B, and C. This time, lets say A is conditionally independent of C given B. Rather than writing out

$$P(A, B, C) = P(A|BC)P(B|C)P(C)$$

We can instead write

$$P(A, B, C) = P(A|B)P(B|C)P(C)$$

While this only removed one variable from a conditional independence, it can lead to massive speed ups in computation time. For example, Consider the distribution over binary random variables B, C, D, E, F, G, H. There are 2^7 possibilities, so the truth table for $P(A|B, C, D, E, F, G, H)$ is quite large. If A was conditionally independent of C, D, E, F, G, H given B, then we could simplify this to $P(A|B)$. Now there only two possibilities, B or $\neg B$.

2 Bayesian Network

2.1 Explanation

A Bayesian Network, typically called a Bayes Net, is a directed acyclic graph in which each edge corresponds to a conditional dependence and each node corresponds to a unique random variable.

For example consider:

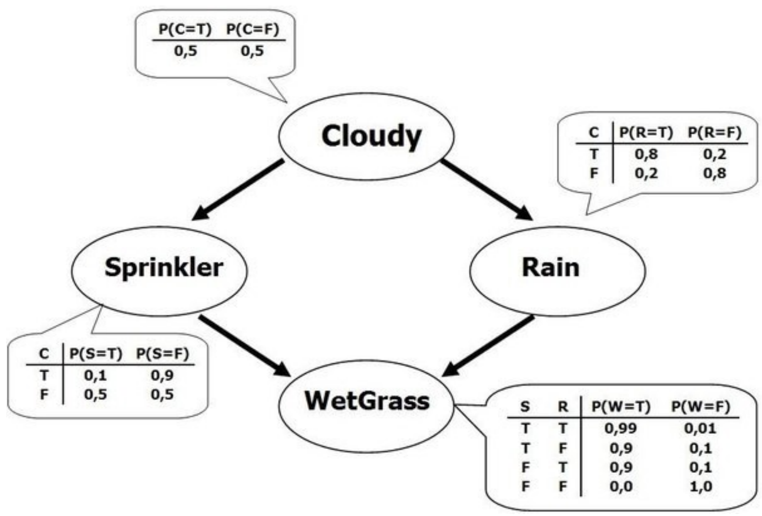

Figure 1: Source: https://towardsdatascience.com/introduction-to-bayesian-networks-81031eeed94e

Cloudy has no parents in the graph, so it is not conditionally dependent on anything. Sprinkler has one parent, Cloudy, so we write its distribution like so: $P(S|C)$. The same applies for Rain. Lastly, WetGrass has two parents, so we write its probability distribution as being conditioned over S and R like so: $P(WG|S,R)$.

2.2 Importance of Directed Edges

Note, Bayesian Networks use directed edges. In other words, an edge from A to B does not imply an edge from B to A. For this reason, when an edge A to B exists, we write $P(B|...A)$, but not $P(A|...B)$.

2.3 Written More Generally

Written more generally, we say a node in a Bayes Net is conditionally independent of its non-descendants given its parents. In other words, WetGrass is independent of Cloudy given Rain and Sprinkler.

2.4 Building a Bayes Net: Importance of Ordering

Let's say you are going to build a Bayes Net. We can think about this in the context of chain rule.

You incrementally add each node to the graph one by one. Each time you add a node, we could condition it on all previously added nodes like chain rule states (add a directed edge for each conditional dependence). But as we learned before, this is unnecessary given there may be conditional independencies. Instead, add the minimum number of edges needed to satisfy the property that the node is conditional independent of all other nodes in the graph given its parents. Repeat until all nodes are added.

Does the order in which you add nodes matter? Yes. Consider the case where You have random variables A, B, C, D. A has no conditional independencies with the other variables. If you had to choose, would you rather to add it first or last? First, otherwise you would have to draw an edge from all other nodes to it. What about if B is conditionally independent of C, D given A. Would you want to add it first or last if you had to choose? Last isn't a bad choice here, as it only depends on A, so number of edges is kept small.

Students can try defining some random variables, with conditional independencies, to further see how ordering matters.

3 Example

You ordered some food and your friends are coming over to watch March Madness. Assume your friends live close enough that they are waking over. The variables we are tracking are

$Friends, DeliveryPerson, CarInDriveway, ChatteringOutside, DoorbellRang$

We come up with the following Bayes Net:

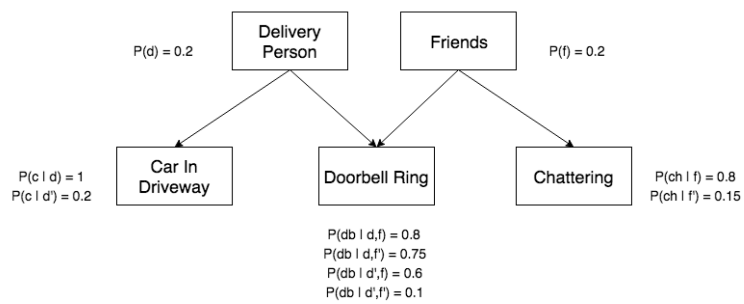

Figure 2: Example Net

If there's a car in the driveway, you may be inclined to believe that the delivery person is here (remember your friends walk). If you hear chattering outside, perhaps you reach the opposite conclusion. What about both? Lets write out the symbolic probability expressions for this Bayes Net.

3.1 Write How Many and Which Variables Each Node Depends On

For example: Delivery Person depends on nothing, it has no parents.

Answer:

$P(DeliveryPerson)$
$P(Friends)$
$P(Car|Delivery)$
$P(Doorbell|Delivery, Friends)$
$P(Chattering|Friends)$

3.2 How Would you Go About Computing An Event?

For example, lets say you wanted to know the probability of doorbell, delivery, not friends, car, and not chattering. In other words:

$$P(Doorbell, Delivery, \neg Friends, Car, \neg Chattering)$$

Note, this has not yet been gone over in class, so you can gauge interest and proceed if people are interested. Make it more of a thought experiment. Answer provided below.

Answer:

Using Db = Doorbell, D = Delivery, F = Friends, Ch = Chattering, and C = Car:

$$\begin{aligned}
P(Db, D, \neg F, C, \neg Ch) &= P(Db|D, \neg F)P(D, \neg F, C, \neg Ch) \\
&= P(Db|D, \neg F)P(C|D)P(D, \neg F, \neg Ch) \\
&= P(Db|D, \neg F)P(C|D)P(\neg Ch|\neg F)P(D, \neg F) \\
&= P(Db|D, \neg F)P(C|D)P(\neg Ch|\neg F)P(D)P(\neg F)
\end{aligned}$$