COMPSCI 240: Reasoning Under Uncertainty

Andrew Lan and Nic Herndon

University of Massachusetts at Amherst

Spring 2019

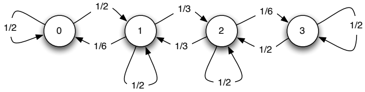
Lecture 30: Bayesian Networks

Bayesian Networks

- We use a Directed Acyclic Graph (DAG) to encode conditional independence assumptions.
  - Nodes X_i in the graph G represent random variables.
  - A directed edge $X_j \rightarrow X_i$ means X_i directly depends on X_j (not causation!).
  - We also define that X_j is a "parent" of X_i.
  - The set of variables that are parents of X_i is denoted Pa_i.
  - X_i is independent of all its nondescendants given Pa_i.
  - The factor associated with variable X_i is $P(X_i|Pa_i)$.

Bayesian Networks vs. Markov Chains

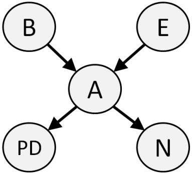
- In Transition Probability Graphs of Markov Chains, **nodes** represent all possible **states**, and **arrows** represents the **probability of transition** from one state to another (with numbers written on it).
- In Bayesian Networks, **nodes** represent all possible **random variables**, and **arrows** represents **dependencies** between the random variables (no numbers associated with it).



P(B=1)
0.001

P(E=1)
0.002

B	E	P(A=1|B,E)
1	1	0.950
1	0	0.940
0	1	0.290
0	0	0.001



A	P(PD=1|A)
1	0.900
0	0.005

A	P(N=1|A)
1	0.750
0	0.100

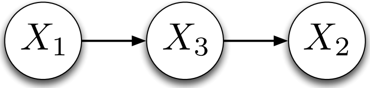
The Bayesian Network Theorem

- **Definition:** A joint PMF $P(X_1, ..., X_d)$ is a Bayesian network with respect to a directed acyclic graph G with parent sets $\{Pa_1, ..., Pa_d\}$ if and only if:

$$P(X_1, ..., X_d) = \prod_{i=1}^{d} P(X_i | Pa_i)$$

- In other words, to be a valid Bayesian network for a given graph G, the joint PMF must factorize according to G.

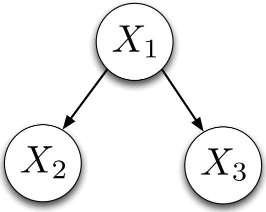
3 Cases of Conditional Independence to Remember



$$Pa_1 = \{\}, Pa_3 = \{X_1\}, Pa_2 = \{X_3\}$$

$$P(X_1, X_2, X_3) = P(X_1)P(X_3|X_1)P(X_2|X_3)$$

3 Cases of Conditional Independence to Remember



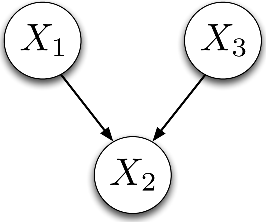
$$Pa_1 = \{\}, Pa_3 = \{X_1\}, Pa_2 = \{X_1\}$$
$$P(X_1, X_2, X_3) = P(X_1)P(X_3|X_1)P(X_2|X_1) \tag{1}$$

- Note that X_2 and X_3 are conditionally independent given X_1:

$$P(X_2, X_3|X_1) = P(X_2|X_1) \cdot P(X_3|X_1)$$

Proof: divide both sides in (1) by $P(X_1)$

3 Cases of Conditional Independence to Remember



$$Pa_1 = \{\}, Pa_3 = \{\}, Pa_2 = \{X_1, X_3\}$$

$$P(X_1, X_2, X_3) = P(X_1)P(X_3)P(X_2|X_1, X_3) \qquad (2)$$

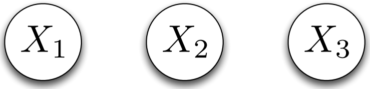
- Note that X_1 is not independent of X_3 given X_2:

$$P(X_1, X_3|X_2) \neq P(X_1|X_2) \cdot P(X_3|X_2)$$

Proof: divide both sides in (2) by $P(X_2)$:

$$P(X_1, X_3|X_2) = \frac{P(X_1)P(X_3)P(X_2|X_1, X_3)}{P(X_2)} \neq P(X_1|X_2) \cdot P(X_3|X_2)$$

If All Nodes Are Independent



$$Pa_1 = \{\}, Pa_3 = \{\}, Pa_2 = \{\}$$

$$P(X_1, X_2, X_3) = P(X_1)P(X_3)P(X_2)$$

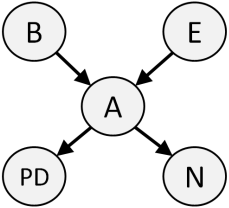
The Alarm Network: Random Variable

- You live in quiet neighborhood in the suburbs of LA. There are two reasons the alarm system in your house will go off: your house is broken into or there is an earthquake. If your alarm goes off you might get a call from the police department. You might also get a call from your neighbor.

P(B=1)
0.001

P(E=1)
0.002

B	E	P(A=1|B,E)
1	1	0.950
1	0	0.940
0	1	0.290
0	0	0.001



A	P(PD=1|A)
1	0.900
0	0.005

A	P(N=1|A)
1	0.750
0	0.100

$$P(B,E,A,PD,N) = P(B)P(E)P(A|B,E)P(PD|A)P(N|A)$$

The Alarm Network: Marginal Query

- **Question:** What is the probability that there was a break-in, but no earthquake, the police call, but your neighbor does not call?

$$
\begin{aligned}
& P(B=1, E=0, PD=1, N=0) \\
&= \sum_{A=\{0,1\}} P(B=1, E=0, PD=1, N=0, A) \\
&= P(B=1, E=0, PD=1, N=0, A=0) \\
&+ P(B=1, E=0, PD=1, N=0, A=1) \\
&= P(B=1)P(E=0)P(A=1|B=1, E=0)P(PD=1|A=1)P(N=0|A=1) \\
&+ P(B=1)P(E=0)P(A=0|B=1, E=0)P(PD=1|A=0)P(N=0|A=0)
\end{aligned}
$$

$$
\begin{aligned}
&= 0.001 \cdot (1-0.002) \cdot 0.94 \cdot 0.9 \cdot (1-0.75) \\
&+ 0.001 \cdot (1-0.002) \cdot (1-0.94) \cdot 0.005 \cdot (1-0.1) = 0.00021\ldots
\end{aligned}
$$

The Alarm Network: Marginal Query

- **Question:** What is the probability that the alarm will be on?

$$
\begin{aligned}
&P(A=1)\\
&=\sum_B\sum_E\sum_{PD}\sum_N P(A=1,B,E,PD,N)\\
&=\sum_B\sum_E\sum_{PD}\sum_N P(B)P(E)P(A=1|B,E)P(PD|A=1)P(N|A=1)\\
&=P(B=0)P(E=0)P(A=1|B,E=0)P(PD=0|A=1)P(N=0|A=1)\\
&+P(B=0)P(E=0)P(A=1|B,E=0)P(PD=0|A=1)P(N=1|A=1)\\
&\cdots\\
&+P(B=1)P(E=1)P(A=1|B,E=1)P(PD=1|A=1)P(N=1|A=1)
\end{aligned}
$$

The Alarm Network: Marginal Query

- We can compute the above using a simple algorithm:

```
Z = 0;
for B = 0 to 1 do
    for E = 0 to 1 do
        for PD = 0 to 1 do
            for N = 0 to 1 do
                Z = Z+
                  P(B)P(E)P(A = 1|B, E)P(PD|A = 1)P(N|A = 1);
            end
        end
    end
end
```

- What would be the potential problem with this?
  - Computational complexity explodes as # of variables increases
  - The multiplication of small number approaches to 0 as # of variables increases

The Alarm Network: Marginal Query

- We can optimize the computation as the following

$$
\begin{aligned}
&P(A=1) \\
&= \sum_{B}\sum_{E}\sum_{PD}\sum_{N} P(B)P(E)P(A=1|B,E)P(PD|A=1)P(N|A=1) \\
&= \sum_{B}\sum_{E} P(B)P(E)P(A=1|B,E)\sum_{PD} P(PD|A=1)\sum_{N} P(N|A=1) \\
&= \sum_{B}\sum_{E} P(B)P(E)P(A=1|B,E)
\end{aligned}
$$

The Alarm Network: Conditional Query

- **Question:** What is the probability that the alarm went off **given that** there was a break-in, but no earthquake, the police call, but your neighbor does not call?

$$P(A = 1|B = 1, E = 0, PD = 1, N = 0)$$

$$= \frac{P(B = 1, E = 0, A = 1, PD = 1, N = 0)}{P(B = 1, E = 0, PD = 1, N = 0)}$$

$$= \frac{P(B = 1, E = 0, A = 1, PD = 1, N = 0)}{\sum_{a=0}^{1} P(B = 1, E = 0, A = a, PD = 1, N = 0)}$$

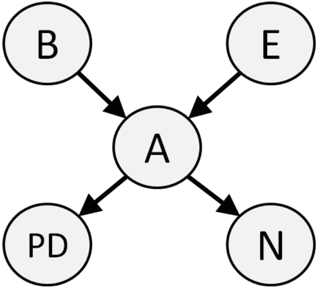
$$= \frac{P(B = 1)P(E = 0)P(A = 1|B = 1, E = 0)P(PD = 1|A = 1)P(N = 0|A = 1)}{\sum_{a=0}^{1} P(B = 1)P(E = 0)P(A = a|B = 1, E = 0)P(PD = 1|A = a)P(N = 0|A = a)}$$

Answering Probabilistic Queries

- **Joint Query:** To compute the probability of an assignment to all of the variables we simply express the joint probability as a product over the individual factors. We then look up the correct entries in the factor tables and multiply them together.
- **Marginal Query:** To compute the probability of an observed subset of the variables in the Bayesian network, we sum the joint probability of all the variables over the possible configurations of the unobserved variables.
- **Conditional Query:** To compute the probability of one subset of the variables given another subset, we first apply the conditional probability formula and then compute the ratio of the resulting marginal probabilities.

Estimating Bayesian Networks from Data

- Just as with simpler models like the biased coin, we can estimate the unknown model parameters from data.



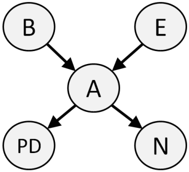
- If we have data consisting of *n* observations of all of the variables in the network, we can easily estimate the entries of each conditional probability table.

Estimating Bayesian Networks: Counting

- **No Parents:** For a variable X with no parents, the estimate of $P(X = x)$ is just the number of times that the variable X takes the value x in the data, divided by the total number of data cases n.
- **Some Parents:** For a variable X with parents $Y_1, ..., Y_p$, the estimate of $P(X = x | Y_1 = y_1, ..., Y_p = y_p)$ is just the number of times that the variable X takes the value x when the parent variables $Y_1, ..., Y_p$ take the values $y_1, ..., y_p$, divided by the total number of times that the parent variables take the values $y_1, ..., y_p$.

Computing the Factor Tables from Observations

- Suppose we have a sample of data as shown below. Each row i is a joint configuration of all of the random variables in the network.



E	B	A	PD	N
1	0	1	1	1
0	0	0	0	1
0	0	1	1	0
0	1	1	1	0
0	0	0	0	0

- In the alarm network, consider the factor $P(E)$. We need to estimate $P(E = 0)$ and $P(E = 1)$.
- Given our data sample, we get the answers $P(E = 0) = 4/5$ and $P(E = 1) = 1/5$.

Computing the Factor Tables from Observations

- In the alarm network, consider the factor $P(N|A)$. We need to estimate $P(N=0|A=0)$,$P(N=1|A=0)$, $P(N=0|A=1)$, $P(N=1|A=1)$. How can we do this?

E	B	A	PD	N
1	0	1	1	1
0	0	0	0	1
0	0	1	1	0
0	1	1	1	0
0	0	0	0	0

- $P(N=0|A=0)=\frac{1}{2}, P(N=1|A=0)=\frac{1}{2}$
- $P(N=0|A=1)=\frac{2}{3}, P(N=1|A=1)=\frac{1}{3}$

Learning the structure of a Bayesian Network from data

- What if you have a dataset, but you do not know the dependencies that exist between the random variables?
- In other words, you do not know what is the graph of your Bayesian network.
- You can estimate the structure of the graph from data!