EECS 496: Sequential Decision Making

Soumya Ray

sray@case.edu

Office: Olin 516

Office hours: T 4-5:30 or by appointment

Recap

- To make probabilistic inference practical on a large scale, there are two key ideas. First, in most scenarios, most random variables ____________.
- Second, given the above, a joint pdf can be represented as a ______. Then the _____ of the ____ can be used to make the inference ______.
- These ideas lead to _______ _______.
- A BN is a way to represent the ________ over ________ as a ________.
- The nodes are _____. The edges are ________.
- A BN is a ______. This makes sense because _______.
- However a BN is not _______. This is because _________.
- An arbitrary DAG represent the joint distribution if for all nodes i, _________. This is called the _______________.
- An edge in a BN represents _________ but not ________. Networks constructed to be _______ will typically be _____, but general ____ are not _____.
- What is explaining away?

Today

- Inference in Bayesian Networks (Ch 14, Russell and Norvig)

BNA and the Chain rule

- So for an arbitrary DAG,

$$\Pr(x_1, \ldots, x_n)$$

$$= \Pr(x_n) \prod_{i=1}^{n-1} \Pr(x_i \mid \{x_j\}_{j=i+1}^{n})$$

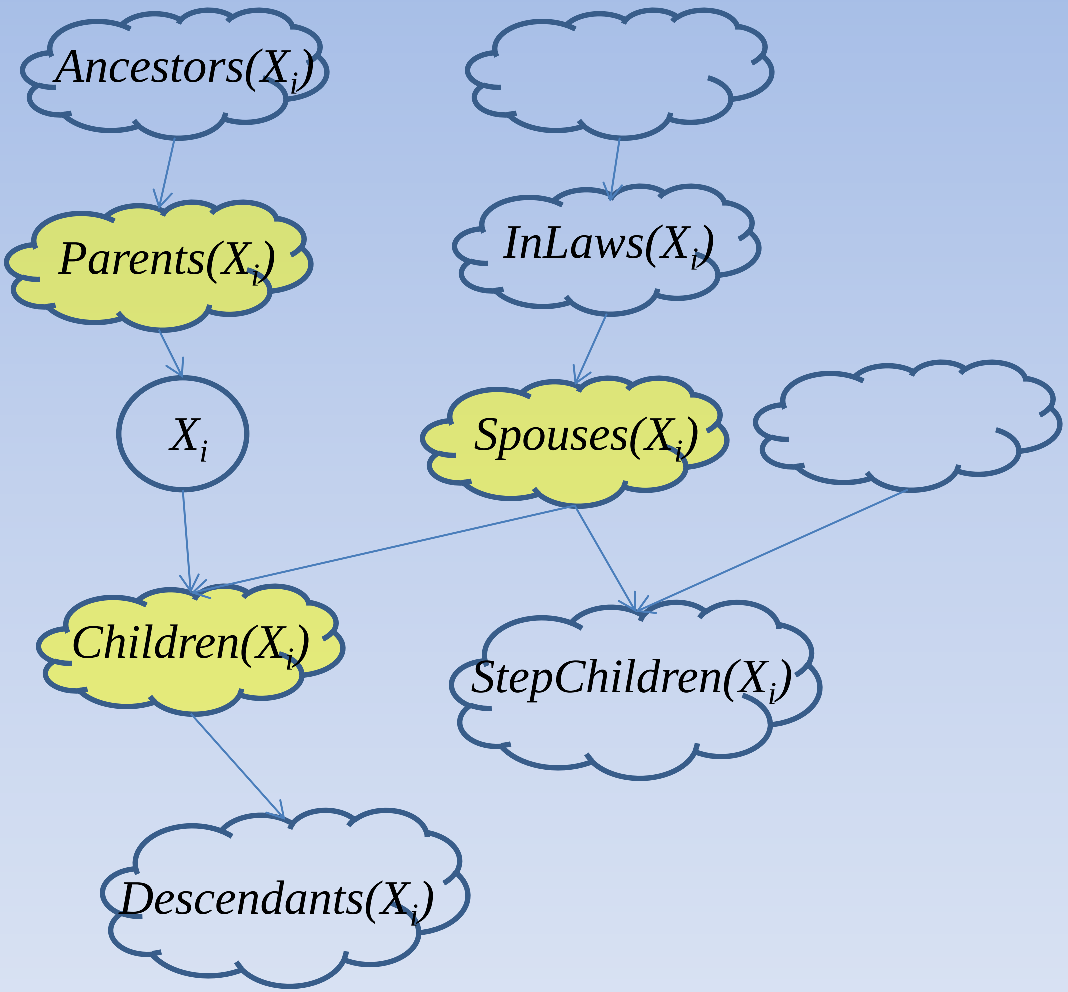
$$\boxed{= \prod_{i=1}^{n} \Pr(x_i \mid Pa(x_i)) \qquad \text{By BNA}}$$

The Markov Blanket

- The Markov Blanket of a node is defined as:
  - Its parents
  - Its children
  - Its children's other parents

- Result: A node in a BN is conditionally independent of *all other nodes* given its Markov Blanket

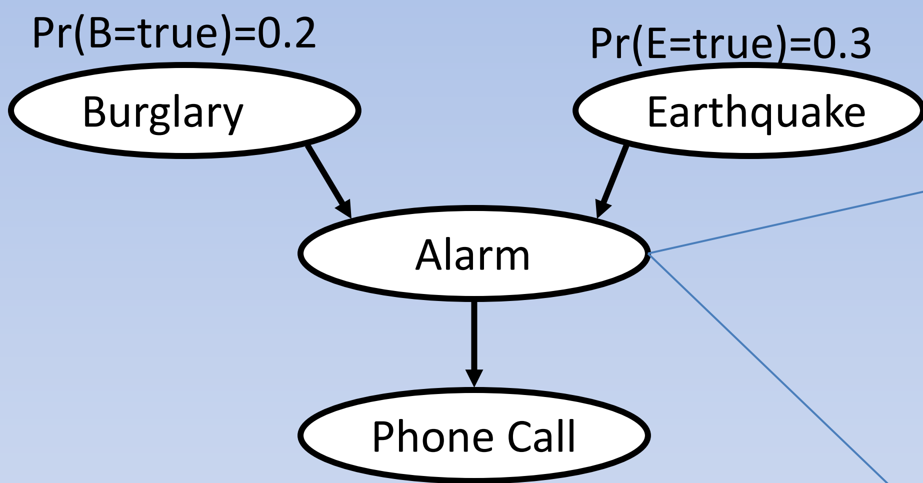
Markov Blanket



Representing Probabilities

- How to represent $Pr(x_i|Pa(x_i))$?
  - (Assume all the r.v.'s are discrete)
- Often represented as a table, the "**conditional probability table**"
- For each value of x_i and $Pa(x_i)$, write down the probability

Conditional Probability Tables



Burglary	Earthquake	A=true
False	False	0.2
False	True	0.5
True	False	0.75
True	True	0.9

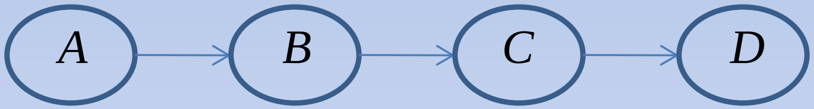
Alarm	P=true
False	0.25
True	0.75

Inference in Bayesian Networks

- How to answer queries $\Pr(V=v \mid \mathbf{E}=\mathbf{e})$, *given a BN*?
- Two kinds of algorithms:
  - Exact
    - Always returns exact answer, but may take a long time
  - Approximate
    - Returns approximate answer. More time=better answers ("anytime")

Variable Elimination

- Suppose we had the BN:



- And we want $Pr(D)$

Variable Elimination

$$\Pr(D) = \sum_{A,B,C} \Pr(A,B,C,D)$$ ← Inference by enumeration

$$= \sum_{A,B,C} \Pr(A)\Pr(B \mid A)\Pr(C \mid B)\Pr(D \mid C)$$ ← By BNA

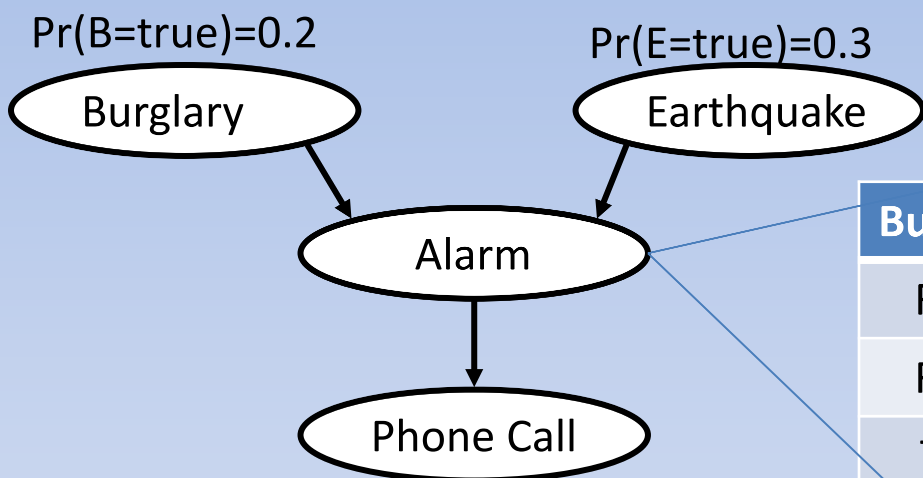
$$= \sum_{C} \Pr(D \mid C) \sum_{B} \Pr(C \mid B) \sum_{A} \Pr(B \mid A)\Pr(A)$$

Each term here is a table, called a "factor". A factor may not be a probability distribution (though in this case it is). Notice that factors are computed by eliminating variables. The efficiency of VE comes from "pushing in" the sums as far as possible.

Variable Elimination

- **Order the variables** in the network with the variable(s) in the query coming last
- For each elimination variable in the ordering
  - **Multiply the tables** involving this variable
  - Then **sum out** this variable by adding all the rows where this variable is the only one changing and the others are fixed
  - Store the resulting “factor” or “potential”

Variable Elimination



Burglary	Earthquake	A=true	A=false
False	False	0.2	0.8
False	True	0.5	0.5
True	False	0.75	0.25
True	True	0.9	0.1

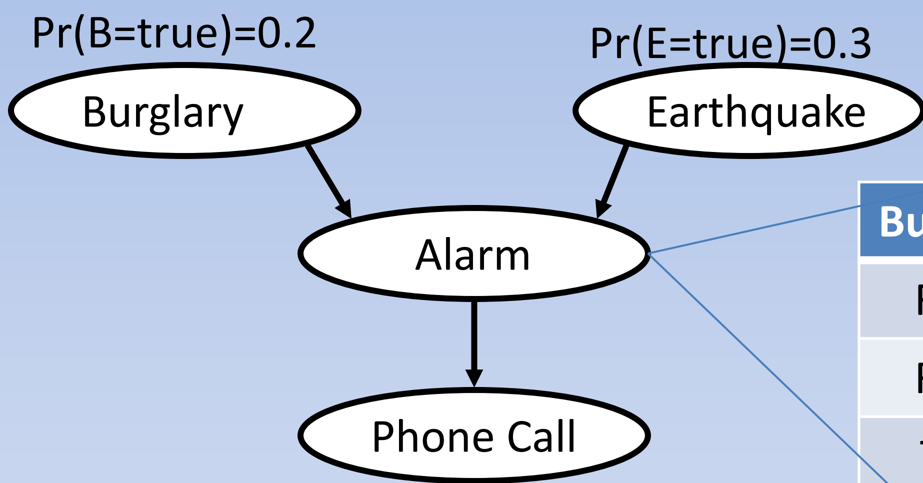
Alarm	P=true	P=false
False	0.25	0.75
True	0.75	0.25

Find $Pr(P)$ using VE

Incorporating evidence

- If we know the value of a variable, just select that value instead of summing out

Variable Elimination



Burglary	Earthquake	A=true	A=false
False	False	0.2	0.8
False	True	0.5	0.5
True	False	0.75	0.25
True	True	0.9	0.1

Alarm	P=true	P=false
False	0.25	0.75
True	0.75	0.25

Find $Pr(P|B=true)$ using VE

Variable Elimination

- The efficiency of this procedure depends on the order of the variables
  - Finding an optimal order is NP-complete
  - However there are good heuristics to choosing a reasonable ordering

Approximate Inference

- Sometimes a BN can be very complex
- Sometimes we don't really need the exact probabilities

- In these cases, we can use *sampling* methods to answer queries
  - Often very fast, very easy to implement
  - But convergence is only asymptotic in general

Approximate Inference

CloudyTomorrow	RainTomorrow	WetGrass	Probability
No	No	No	
No	No	Yes	
No	Yes	No	
No	Yes	Yes	
Yes	No	No	
Yes	No	Yes	
Yes	Yes	No	
Yes	Yes	Yes	

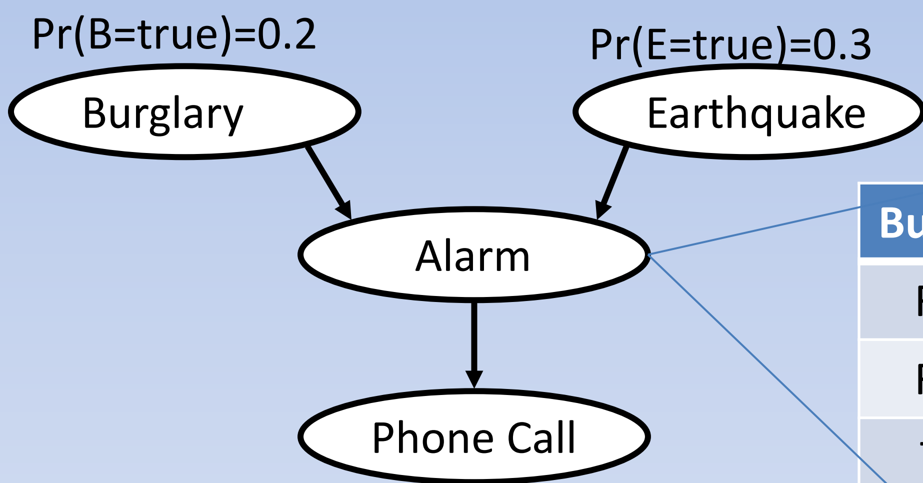
$$\Pr(CloudyTomorrow = Yes)?$$

Approximate Inference (Monte Carlo)

- Generate lots of atomic events from the pdf
- Count samples of desired event, divide by total

Approximate Inference 1 (Monte Carlo)

- How to generate a sample from a BN?



Burglary	Earthquake	A=true	A=false
False	False	0.2	0.8
False	True	0.5	0.5
True	False	0.75	0.25
True	True	0.9	0.1

Alarm	P=true	P=false
False	0.25	0.75
True	0.75	0.25

Approximate Inference 1 (Monte Carlo)

- How to generate a sample from a BN?
- Idea: Topologically sort the variables according to the graph structure
- Sample each according to the conditional distribution (well-defined due to the sorting)
- Count the samples with desired values
- Easy!
  - Right?