INFERENCE IN BAYESIAN NETWORKS

Today

- Reading
  - AIMA 14.4 – 14.5
- Goals
  - Recap d-separation
  - Exact inference in BN
  - Approximate inference in BN

Bayesian Networks

- A Bayesian Network is a directed acyclic graph that represents the conditional independencies of a set of random variables
  - Each random variable corresponds to a node
  - A directed edge represents a direct influence
  - The conditional distribution of each node given its parents must be explicitly specified

Representing the joint using a BN

- Given a BN over a set of random variables, the joint distribution can be factor as

$$p(x_1, \ldots, x_N) = \prod_{i=1}^{N} p(x_i | \text{parents}(x_i))$$

Connection patterns and independence

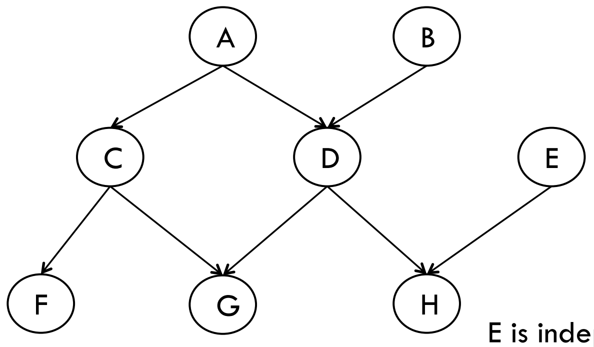
- **Linear connection:** The two end variables are dependent on each other. Observing the middle variable makes them independent.
- **Converging connection:** The two end variables are independent of each other. Observing the middle variable makes them dependent.
- **Divergent connection:** The two end variables are dependent on each other. Observing the middle variable makes them independent.

D-Separation

- Algorithm to determine independencies in BN
- Query: Are two variables X_i and X_j independent?
- Check all paths between X_i and X_j
  - If all paths are blocked, then independent
  - If any path is not blocked then not independent

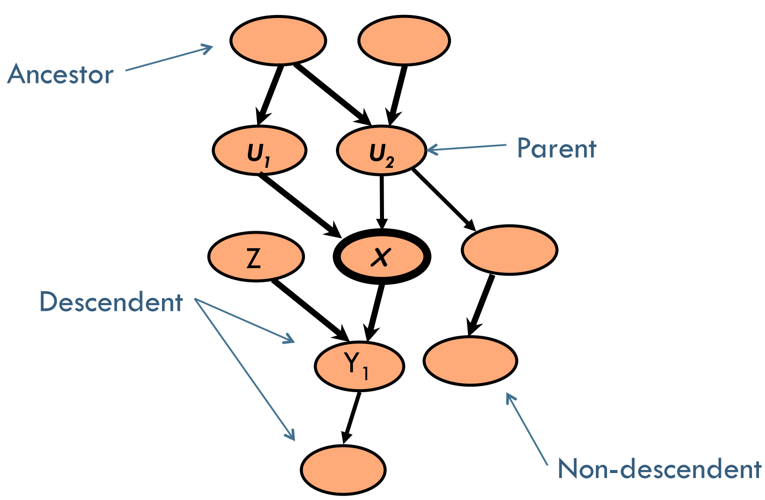
A path is blocked if for any connection on the path the two end variables are independent

List the independencies in the following Bayesian Network

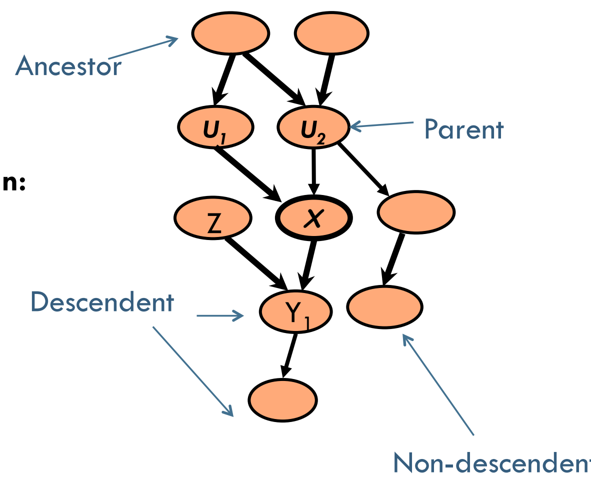


E is independent of G?
C is independent of D?
C is independent of D given G?
F is independent of A given {C,D}?

Bayesian Networks terminology

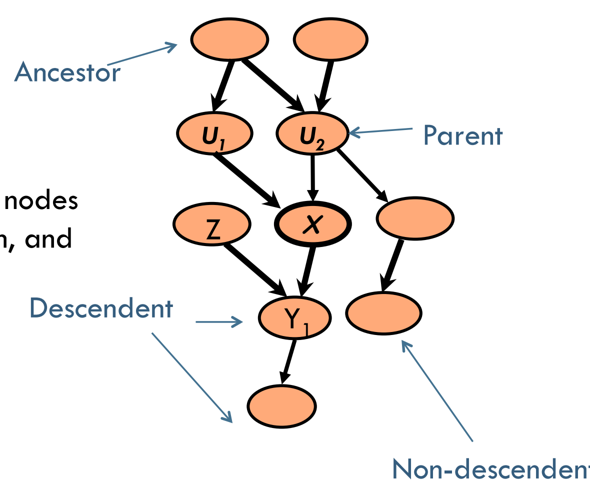


Independence assumptions encoded in the Bayesian Network



Local Markov Assumption:
A node X is independent of its non-descendants given its parents

Independence assumptions encoded in the Bayesian Network

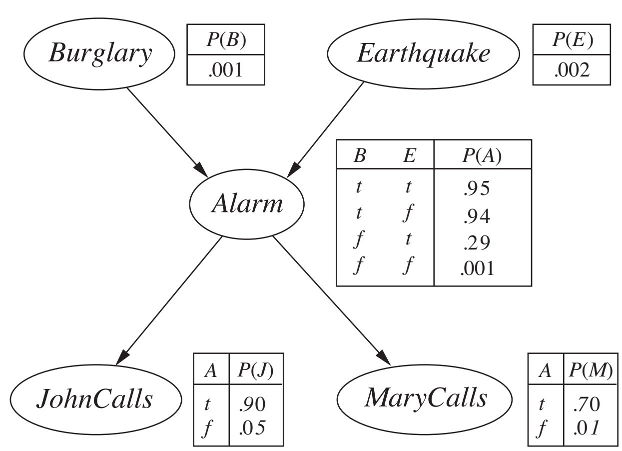


Markov Blanket:
A node X is conditionally independent of all other nodes given its parents, children, and children's parents

Inference in Bayesian Networks

- Probabilistic inference refers to the task of computing some desired probability given other known probabilities (evidence)
- Exact Inference
  - Enumeration
  - Variable elimination
- Approximate Inference
  - Direct sampling
  - Rejection sampling
  - Likelihood weighting
  - MCMC

Recall: Burglary network

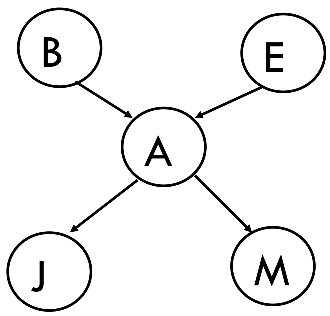


Inference by Enumeration

~~Step One: select the entries in the table consistent with the evidence (this becomes our world)~~

Step Two: sum over the H variables to get the joint distribution of the query and evidence variables

Step Three: Normalize



Practice Queries:

- p(J|E = true)
- p(A|J = true, M=true)
- p(B | E=true, J =true)

Inference by Variable Elimination

- Carry out sums from right to left storing intermediate results to avoid recomputation

$$p(B|j,m) = \alpha\, p(B) \sum_e p(e) \sum_a p(a|B,e)\; p(j|a)\; p(m|a)$$
$$= \alpha\, f_1(B) \sum_e f_2(e) \sum_a f_3(A,B,E)\; f_4(A)\; f_5(A)$$
$$= \alpha\, f_1(B) \sum_e f_2(e)\; f_6(B,E)$$
$$= \alpha\, f_1(B)\; f_7(B)$$

- Results are stored in factors (matrices)
- Two operations: pointwise multiplication and summation

Approximate Inference

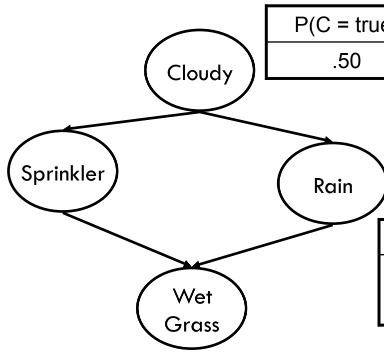
- Analogous to uninformed/informed search algorithms that use an incremental formulation
  - Direct sampling
  - Rejection sampling
  - Likelihood weighting

- Analogous to local search algorithms that use a complete-state formulation and make local modifications
  - Gibbs sampling (special case of MCMC methods)

Lecture proceeds on whiteboard!

Wet Grass Example

P(C = true)
.50



C	P(S=true|C)
T	.10
F	.50

C	P(R=true|C)
T	.80
F	.20

S	R	P(W=true|S,R)
T	T	.99
T	F	.90
F	T	.90
F	F	.01

Gibbs Sampling

- Analogous to a local search algorithm where we make local modifications to our current state
  - Initial state = random assignment of non-evidence variables
  - States = complete assignment of values to variables
  - Transition = sample a new value for each variable in turn

Draw state space for WetGrass example on board

Gibbs Sampling

- Analogous to a local search algorithm where we make local modifications to our current state
  - Initial state = random assignment of non-evidence variables
  - States = complete assignment of values to variables
  - Transition = sample a new value for each variable in turn
- Each step to a new state is recorded as a sample
- In the limit, the probability of being in a state is proportional to that state's posterior probability

Gibbs Sampling

- Gibbs sampling is an instance of a more general class of algorithms known as Markov Chain Monte Carlo (MCMC) algorithms
  - Note the use of the phrase “Markov chain” which we saw an example of earlier
- Other methods you might hear mentioned
  - Metropolis-Hastings (a generalization of Gibbs sampling)
  - Variational method
  - Belief propagation