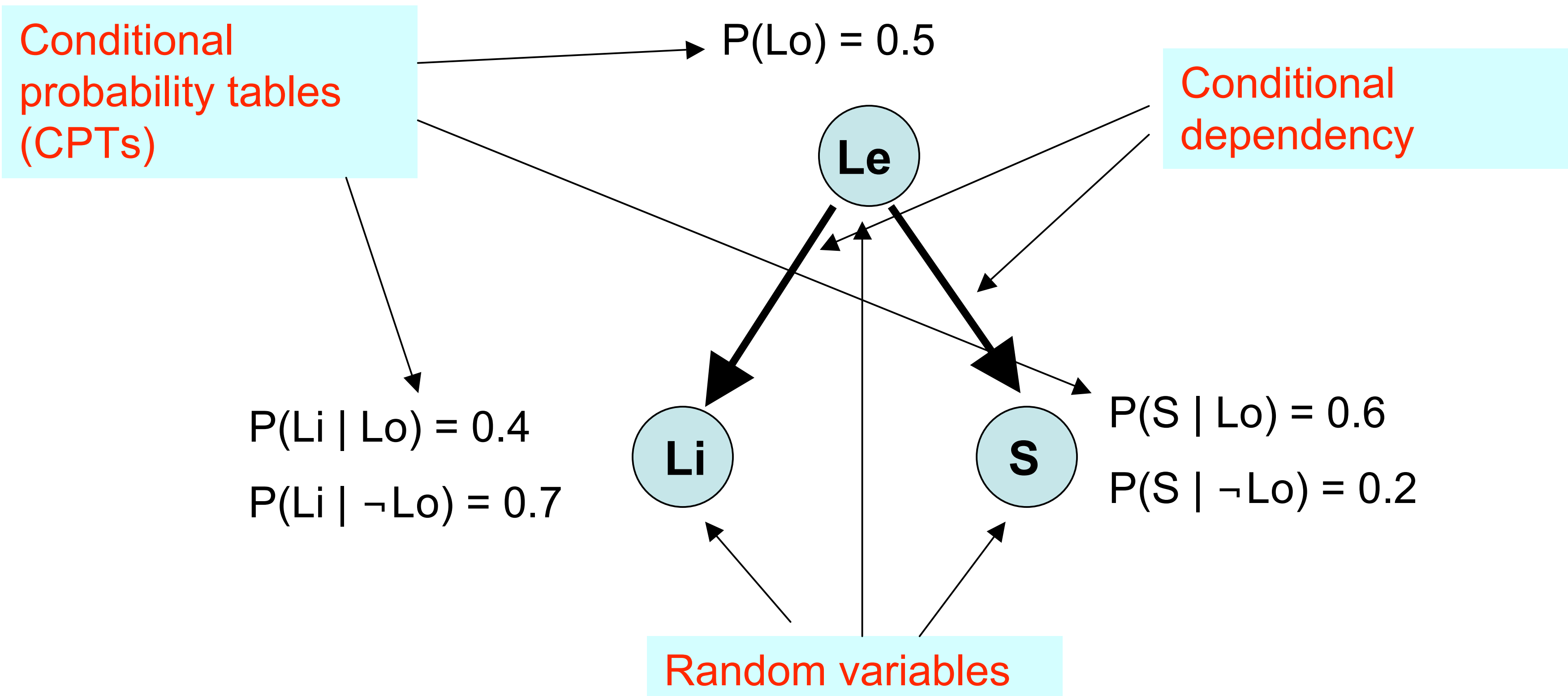
15-780: Graduate Artificial Intelligence

Bayesian networks: Construction and inference

Bayesian networks: Notations

Bayesian networks are directed acyclic graphs.



Constructing a Bayesian network

- How do we go about constructing a network for a specific problem?
- Step 1: Identify the random variables
- Step 2: Determine the conditional dependencies
- Step 3: Populate the CPTs

Can be learned from observation data!

A example problem

- An alarm system

  B – Did a burglary occur?

  E – Did an earthquake occur?

  A – Did the alarm sound off?

  M – Mary calls

  J – John calls

- How do we reconstruct the network for this problem?

Factoring joint distributions

- Using the chain rule we can always factor a joint distribution as follows:

  $P(A,B,E,J,M) =$

  $P(A \mid B,E,J,M)\ P(B,E,J,M) =$

  $P(A \mid B,E,J,M)\ P(B \mid E,J,M)\ P(E,J,M) =$

  $P(A \mid B,E,J,M)\ P(B \mid E, J,M)\ P(E \mid J,M)\ P(J,M)$

  $P(A \mid B,E,J,M)\ P(B \mid E, J,M)\ P(E \mid J,M) P(J \mid M) P(M)$
- This type of conditional dependencies can also be represented graphically.

A Bayesian network

Number of parameters:

A: 2^4

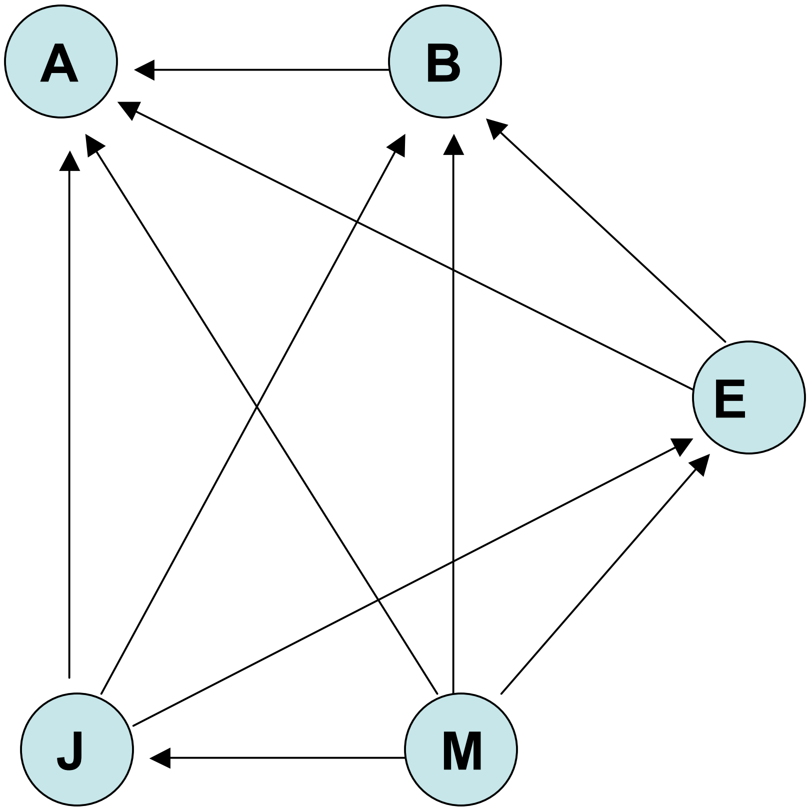
B: 2^3

E: 4

J: 2

M: 1

A total of 31 parameters



A better approach

- An alarm system

  B – Did a burglary occur?

  E – Did an earthquake occur?

  A – Did the alarm sound off?

  M – Mary calls

  J – John calls

- Lets use our knowledge of the domain!

Reconstructing a network

Number of parameters:

A: 4

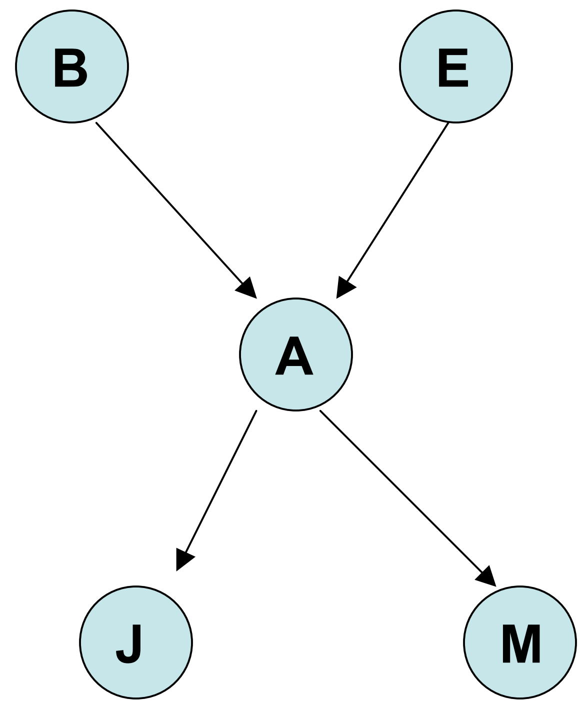
B: 1

E: 1

J: 2

M: 2

A total of 10 parameters



By relying on domain knowledge we saved 21 parameters!

Constructing a Bayesian network: Revisited

- Step 1: Identify the random variables
- Step 2: Determine the conditional dependencies

  - Select on ordering of the variables
  - Add them one at a time
  - For each new variable X added select the minimal subset of nodes as parents such that X is independent from all other nodes in the current network given its parents.

- Step 3: Populate the CPTs

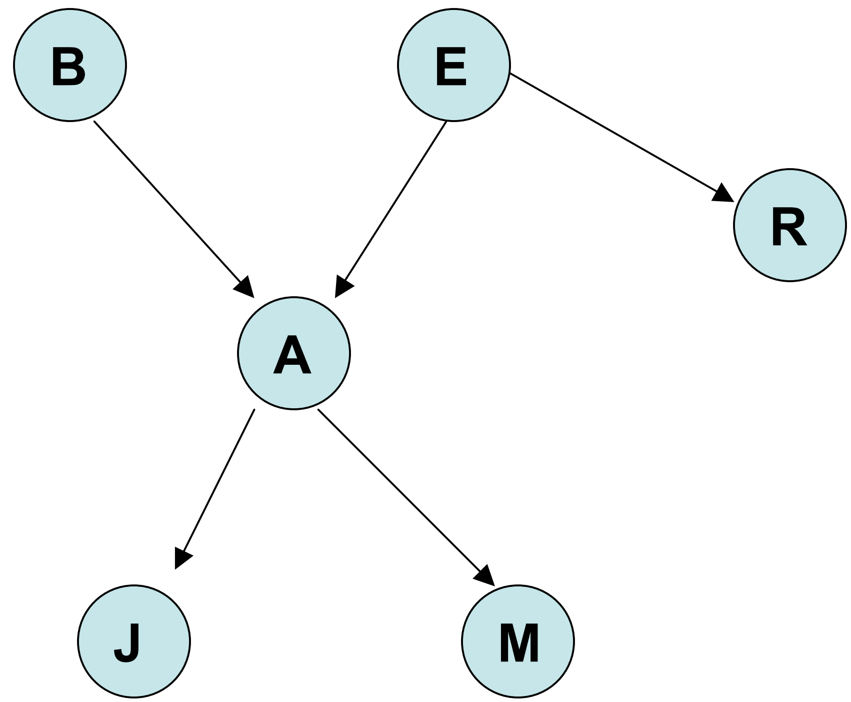
  - We will discuss this when we talk about density estimations

Reconstructing a network

Suppose we wanted to add a new variable to the network:

R – Did the radio announce that there was an earthquake?

How should we insert it?



Bayesian network: Inference

- Once the network is constructed, we can use algorithm for inferring the values of unobserved variables.
- For example, in our previous network the only observed variables are the phone call and the radio announcement. However, what we are really interested in is whether there was a burglary or not.
- How can we determine that?

Inference

- Lets start with a simpler question
  - How can we compute a joint distribution from the network?
  - For example, P(B,¬E,A,J, ¬M)?
- Answer:
  - That’s easy, lets use the network

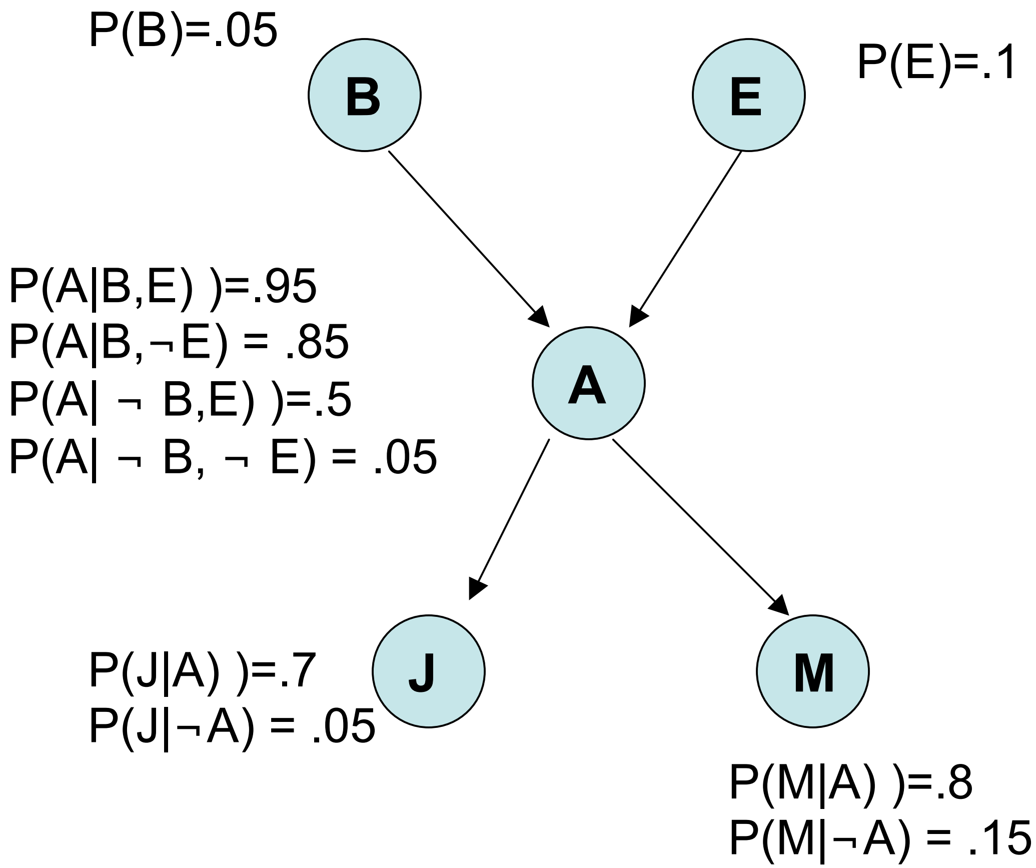
Computing: P(B,¬E,A,J, ¬M)

P(B,¬E,A,J, ¬M) =

P(B)P(¬E)P(A | B, ¬E)
P(J | A)P(¬M | A)

= 0.05*0.9*.85*.7*.2

= 0.005355



Computing: P(B,¬E,A,J, ¬M)

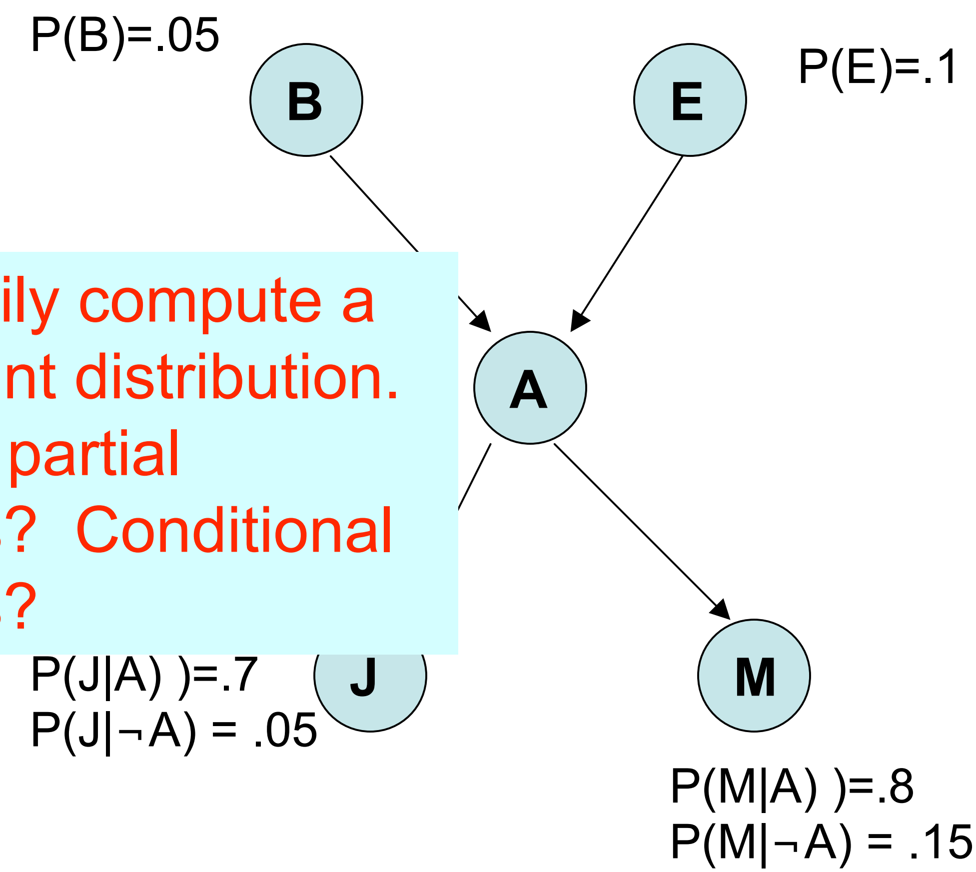
P(B,¬E,A,J, ¬M) =

P(B)P(¬E)P(A | B, ¬E)
P(J | A)P(¬M | A)

= 0.05*0.9*.85*.7*.2

= 0.005355

We can easily compute a complete joint distribution. What about partial distributions? Conditional distributions?

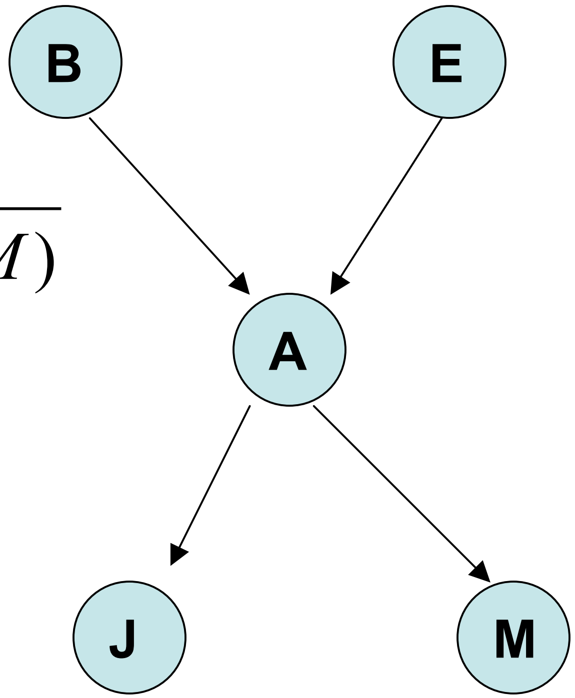


Inference

- We are interested in queries of the form: P(B | J,¬M)
- This can also be written as a joint:

$$P(B \mid J, \neg M) = \frac{P(B, J, \neg M)}{P(B, J, \neg M) + P(\neg B, J, \neg M)}$$

- How do we compute the new joint?



Computing partial joints

$$P(B \mid J, \neg M) = \frac{P(B, J, \neg M)}{P(B, J, \neg M) + P(\neg B, J, \neg M)}$$

Sum all instances with these settings (the sum is over the possible assignments to the other two variables, E and A)

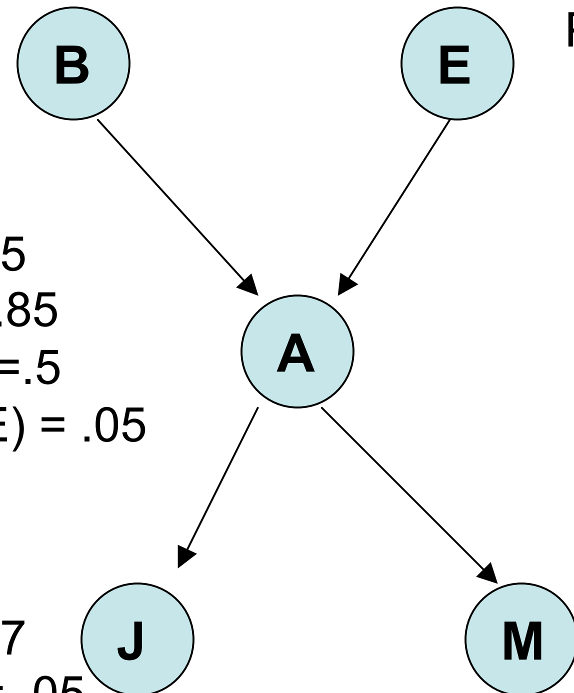
Computing: $P(B,J, \neg M)$

$P(B,J, \neg M) =$

$P(B,J, \neg M,A,E)+$

$P(B,J, \neg M, \neg A,E) +$
$P(B,J, \neg M,A, \neg E) +$
$P(B,J, \neg M, \neg A, \neg E) =$

$0.0007+0.00001+0.005+0.$
$0003 = 0.00601$



Computing partial joints

$$P(B \mid J, \neg M) = \frac{P(B, J, \neg M)}{P(B, J, \neg M) + P(\neg B, J, \neg M)}$$

Sum all instances with these settings (the sum is over the possible assignments to the other two variables, E and A)

- This method can be improved by re-using calculations (similar to dynamic programming)
- Still, the number of possible assignments is exponential in the unobserved variables?
- That is, unfortunately, the best we can do. General querying of Bayesian networks is NP-complete

Inference in Bayesian networks if NP complete (sketch)

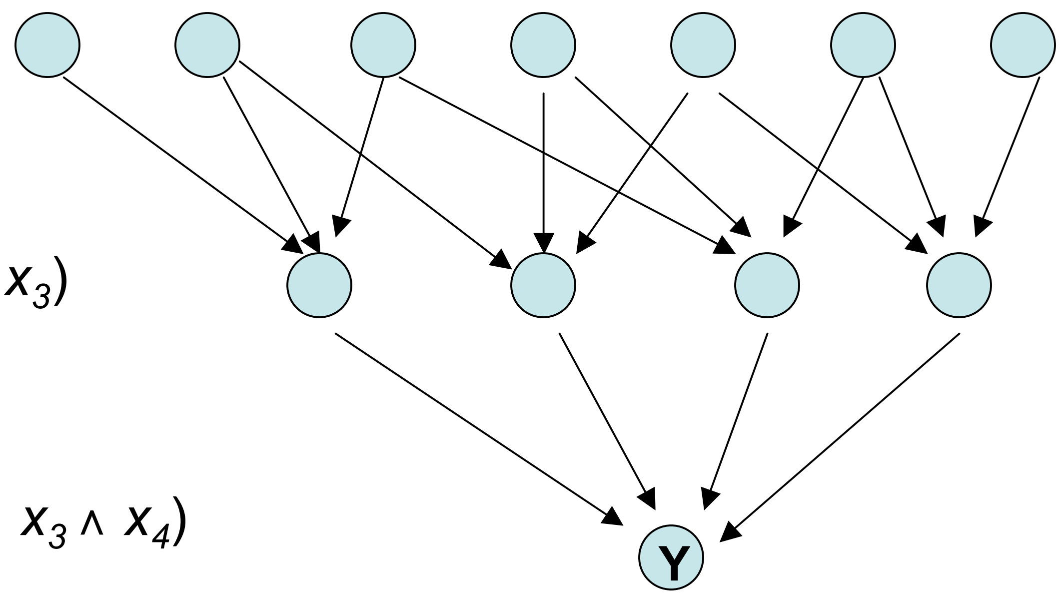
- Reduction from 3SAT
- Recall: 3SAT, find satisfying assignments to the following problem: $(a \vee b \vee c) \wedge (d \vee \neg b \vee \neg c) \ldots$

What is P(Y)?

$P(x_i=1) = 0.5$

$P(x_i=1) = (x_1 \vee x_2 \vee x_3)$

$P(Y=1) = (x_1 \wedge x_2 \wedge x_3 \wedge x_4)$



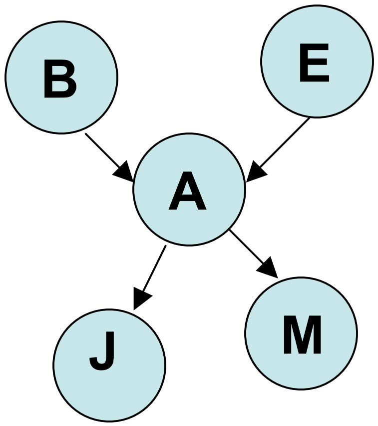
Other inference methods

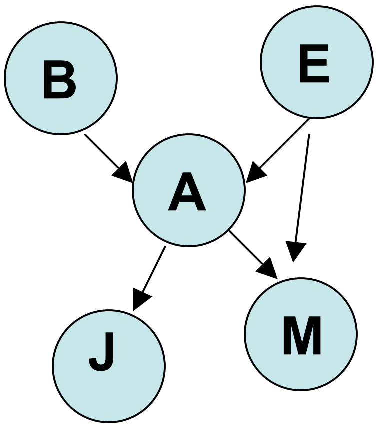
- Convert network to a polytree

  - In a polytree no two nodes have more than one path between them

  - We can convert arbitrary networks to a polytree by clustering (grouping) nodes. For such a graph there is a linear algorithm which is linear in the number of nodes

  - However, converting into a polytree can result in an exponential increase in the size of the CPTs



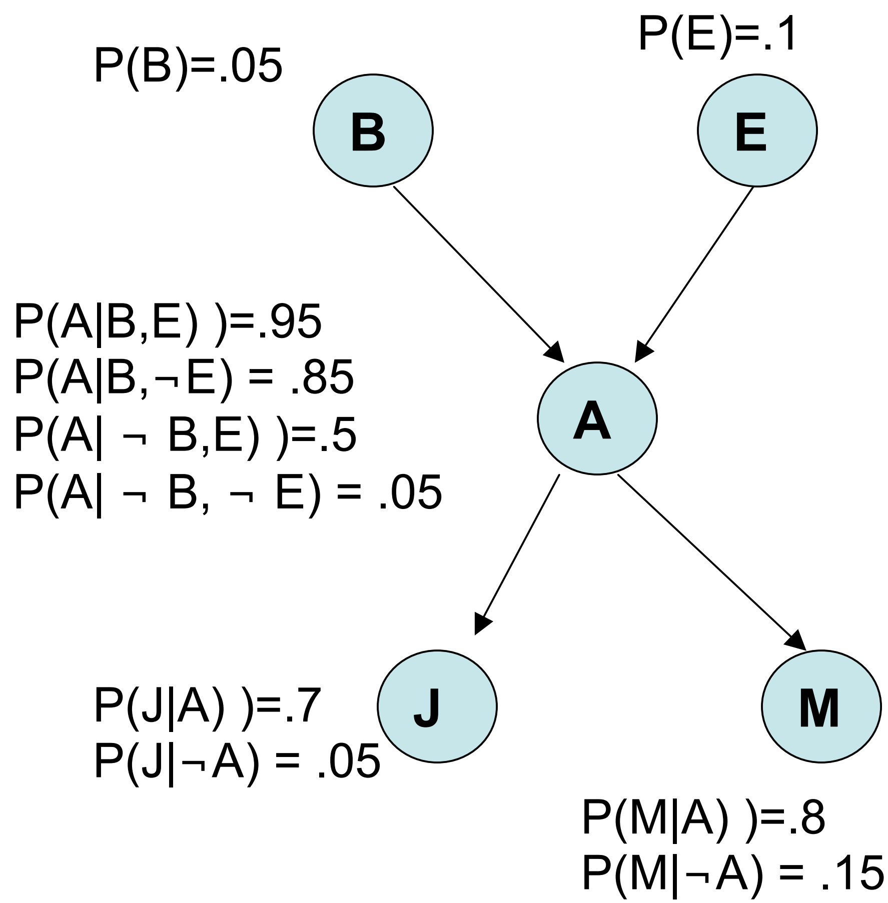


Stochastic inference

- We can easily sample the joint distribution to obtain possible instances

1. Sample the free variable
2. For every other variable:
   - If all parents have been sampled, sample based on conditional distribution

We end up with a new set of assignments for B,E,A,J and M which are a random sample from the joint

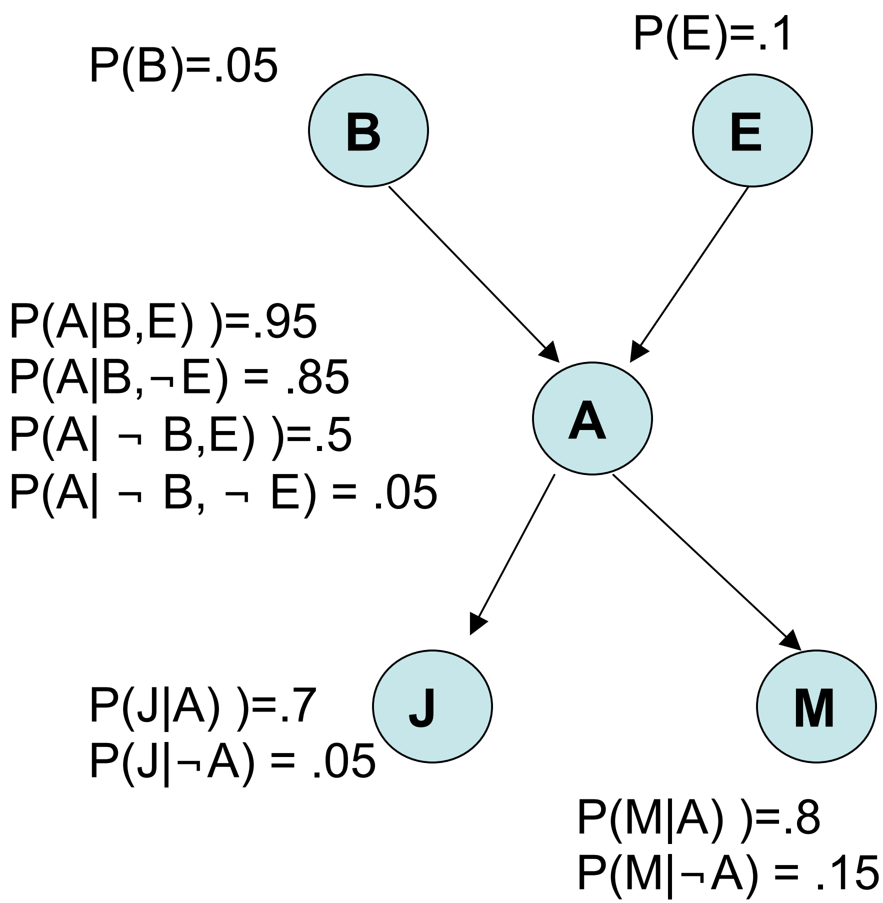


Stochastic inference

- We can easily sample the joint distribution to obtain possible instances

1. Sample the free variable
2. For every other variable:
   - If all parents have been sampled, sample based on conditional distribution

Its always possible to carry out this sampling procedure, why?



Using sampling for inference

- Lets revisit our problem: Compute P(B | J,¬M)
- Looking at the samples we can cound:

  - N: total number of samples
  - N_c : total number of samples in which the condition holds (J,¬M)
  - N_B: total number of samples where the joint is true (B,J,¬M)
- For a large enough N
  - N_c / N ≈ P(J,¬M)
  - N_B / N ≈ P(B,J,¬M)
- And so, we can set

P(B | J,¬M) = P(B,J,¬M) / P(J,¬M) ≈ N_B / N_c

Using sampling for inference

- Lets revisit our problem: Compute P(B | J,¬M)
- Looking at the samples we can cound:

  - *N*: total number o

  - N_c : total number

  - N_B: total number
- For a large enoug

  - N_c / N ≈ P(J,¬M)

  - N_B / N ≈ P(B,J,¬M)
- And so, we can set

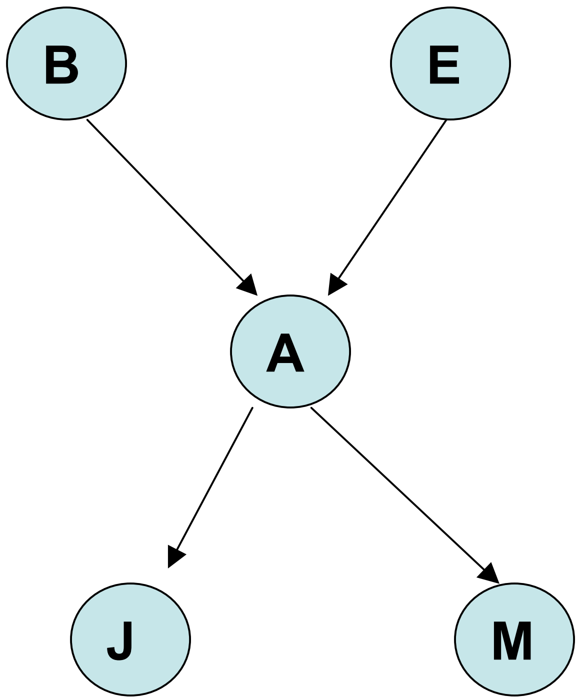
P(B | J,¬M) = P(B,J,¬M) / P(J,¬M) ≈ N_B / N_c

Problem: What if the condition rarely happens?

We would need lots and lots of samples, and most would be wasted

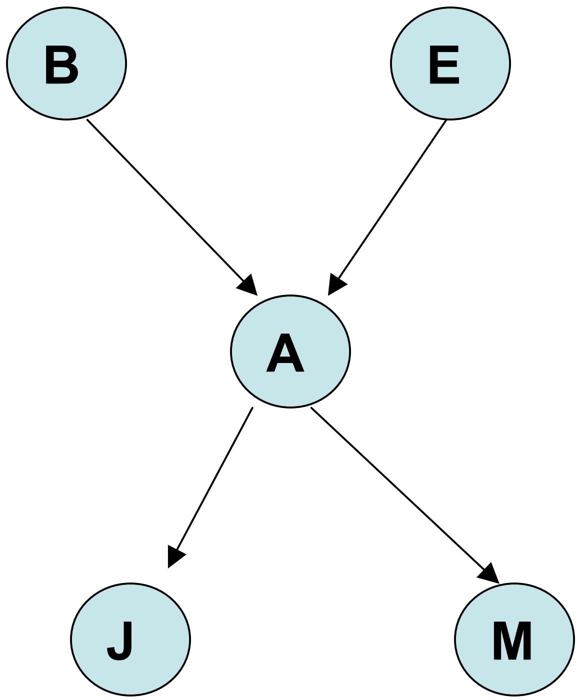
Weighted sampling

- Compute P(B | J,¬M)
- We can manually set the value of J to 1 and M to 0
- This way, all samples will contain the correct values for the conditional variables
- Problems?



Weighted sampling

- Compute P(B | J,¬M)
- Given an assignment to parents, we assign a value of 1 to J and 0 to M.
- We record the *probability* of this assignment ($w = p_1 * p_2$) and we weight the new joint sample by *w*



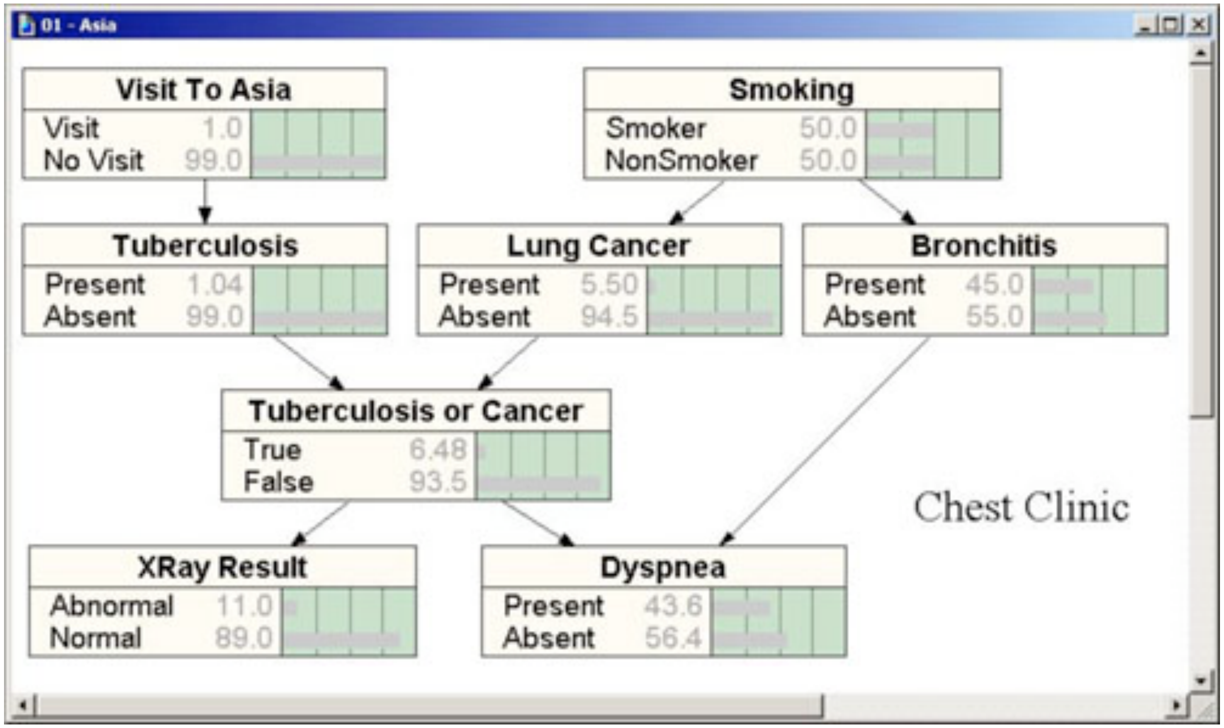
Weighted sampling algorithm for computing P(B | J,¬M)

- Set $N_B, N_c = 0$
- Sample the joint setting the values for *J* and *M*, compute the weight, *w,* of this sample
- $N_c = N_c + w$
- If $B = 1$, $N_B = N_B + w$

- After many iterations, set

P(B | J,¬M) = N_B / N_c

Bayesian networks for cancer detection



Important points

- Bayes rule
- Joint distribution, independence, conditional independence
- Attributes of Bayesian networks
- Constructing a Bayesian network
- Inference in Bayesian networks