Bayes Networks: Representation and Inference

Manuela M. Veloso

Carnegie Mellon University
Computer Science Department

Thanks to past instructors

15-780 Graduate AI – Spring 2013

Readings:

- Russell & Norvig: chapter 14

Bayes Rule

- P (A | B) = P (A , B) / P (B)
- P (A , B) = P (A | B) P (B)
- P (A | B) = P (B | A) P (A) / P (B)

Independence

Liked movie	Slept in movie	Rain	P
1	1	1	0.05
1	0	1	0.1
0	0	1	0.025
0	1	1	0.075
1	1	0	0.15
1	0	0	0.3
0	0	0	0.075
0	1	0	0.225

P(slept) = 0.5

P(slept | rain = 1) = 0.5

Additional knowledge (about Rain) does not change the estimate:

the two random variables - Slept and Rain - are independent.

Independence

Variables R and S are independent if:

P(S | R) = P(S)

And we can then derive:

- P(¬S | R) = P(¬S)
- P(S,R) = P(S)P(R)
- P(R | S) = P(R)

Conditional Independence

Two dependent random variables may become independent when conditioned on a third variable:

$$P(A,B \mid C) = P(A \mid C)\, P(B \mid C)$$

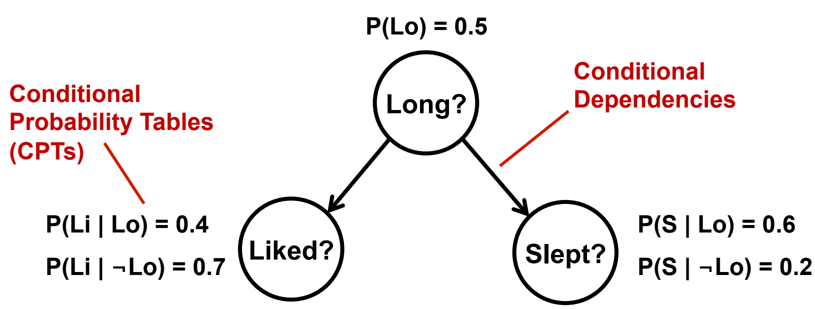
Example:

- P(liked movie) = 0.5
- P(slept) = 0.4
- P(liked movie, slept) = 0.1
- P(liked movie | long) = 0.4
- P(slept | long) = 0.6
- P(like movie, slept | long) = 0.24

Given knowledge of length, the two other variables become independent

Bayesian Networks

Bayesian networks are directed acyclic graphs with nodes representing random variables and edges representing dependency assumptions



Constructing a Bayesian Network

- Step 1: Identify the random variables
- Step 2: Determine the conditional dependencies
- Step 3: Populate the CPTs

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 construct the network for this problem with these 5 random variables?

Factoring Joint Distributions

We can try to get *conditional dependencies of one variable on others (to represent graphically as earlier).*

Using the chain rule we can always *factor* a joint distribution:

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

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

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

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

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

Example: Variables ordered alphabetically. Other orderings fine too.

A Bayesian Network

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

Number of params in CPT:

A: 2^4

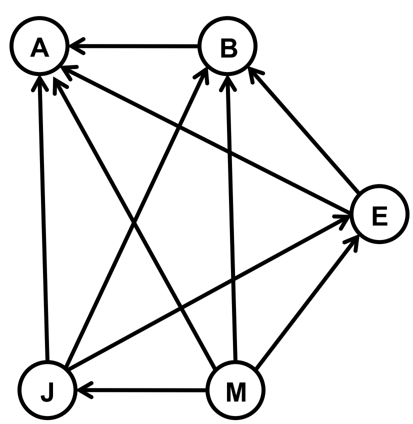
B: 2^3

E: 4

J: 2

M: 1

Total: 31 parameters

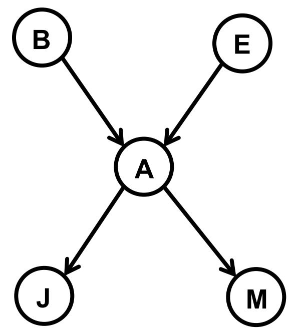


Using Domain Knowledge

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

Domain knowledge: story… (Judea Pearl)



Number of parameters:

A: 4, B: 1, E: 1, J: 2, M: 2

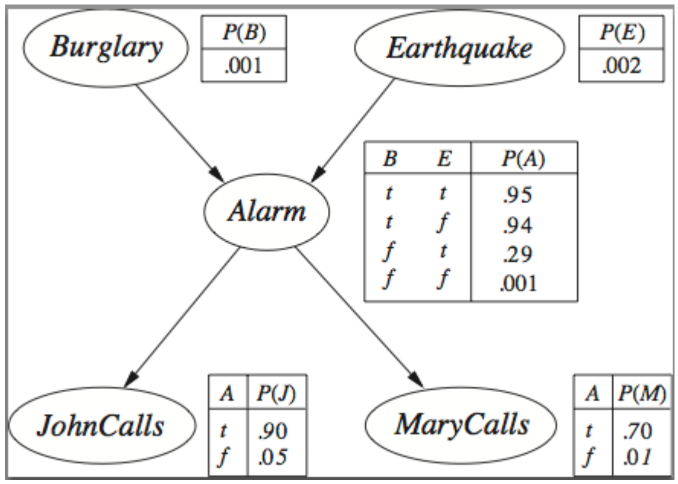
Total: 10 parameters

By using domain knowledge we saved 21 parameters.

Constructing a Bayesian Network: Revisited

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

Bayes Net Example – CPTs text book



Bayesian Network Inference

- Given a network, we want to INFER values for all sorts of unobserved variables or joint variables, or conditional variables – anything!
- For example, in our previous network the only observed variables are the *phone calls* from John and Mary. However, what we are really interested in is the *value of the variable burglary, given our observations.*
- INFERENCE: for example: P (B | J, ¬M)

Inference

A simpler question: How can we compute a joint distribution from the network?

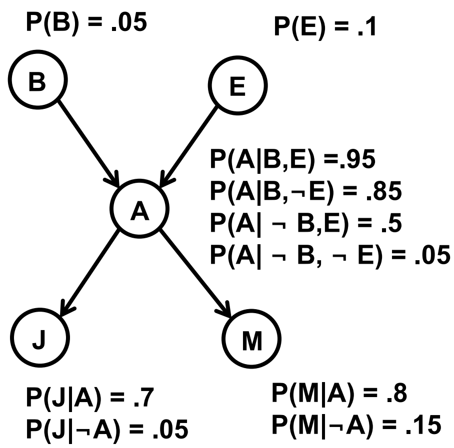
For example, P(B,¬E,A,J, ¬M).

P(B,¬E,A,J, ¬M) =

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

= 0.05*0.9*0.85*0.7*0.2

= 0.005355

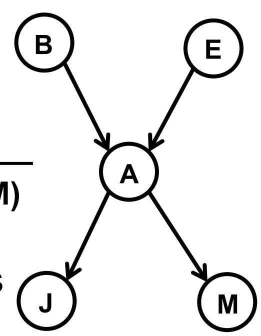


Inference for Partial Joints

We are now interested in queries of the form: P(B | J,¬M)

Conditionals can be written as joints:

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



How do we compute a joint that does not include all the variables of the network (*partial joint*)?

We can “add” the missing variables in all their possible combinations.

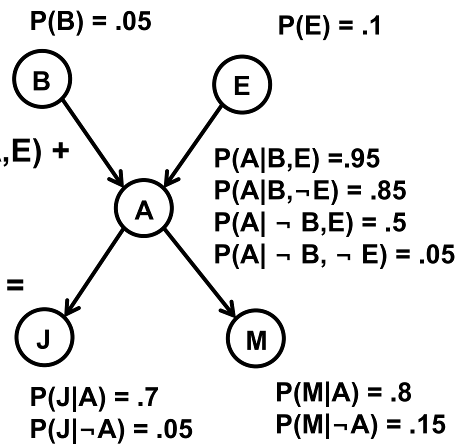
Computing: P(B,J, ¬M)

P(B,J,¬M) =

P(B,J,¬M,A,E) + P(B,J,¬M,¬ A,E) +
P(B,J, ¬M,A,¬E) +
P(B,J,¬M,¬ A,¬E) =

0.0007+0.00001+0.005+0.0003 =

0.00601



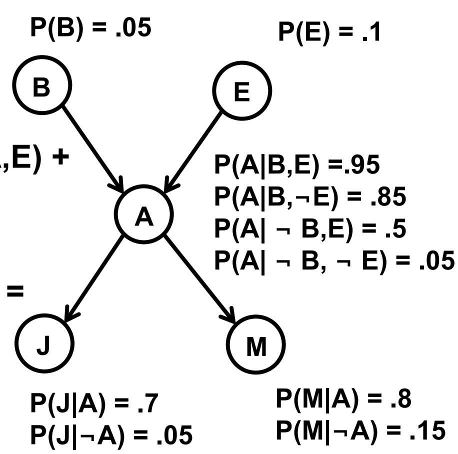
Computing: P(B,J, ¬M)

P(B,J,¬M) =

P(B,J,¬M,A,E) + P(B,J,¬M,¬ A,E) +
P(B,J, ¬M,A,¬E) +
P(B,J,¬M,¬ A,¬E) =

0.0007+0.00001+0.005+0.0003 =

0.00601



How can we reuse computations?

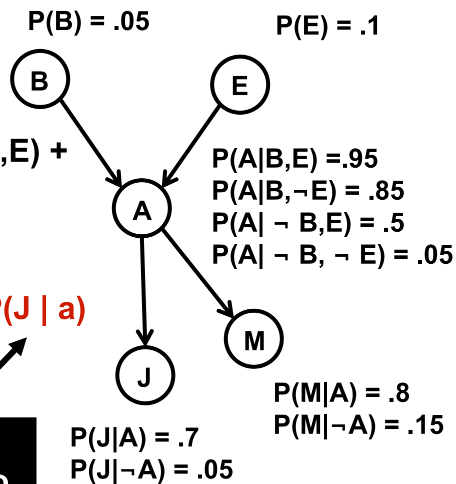
Computing: P(B,J, ¬M)

P(B,J,¬M) =

P(B,J,¬M,A,E) + P(B,J,¬M,¬ A,E) +
P(B,J, ¬M,A,¬E) +
P(B,J,¬M,¬ A,¬E) =

$$\sum_a \sum_e P(B)P(e)P(a \mid B,e)P(M \mid a)P(J \mid a)$$

Store as a function of a and use whenever necessary (no need to recompute each time)



Computational Complexity

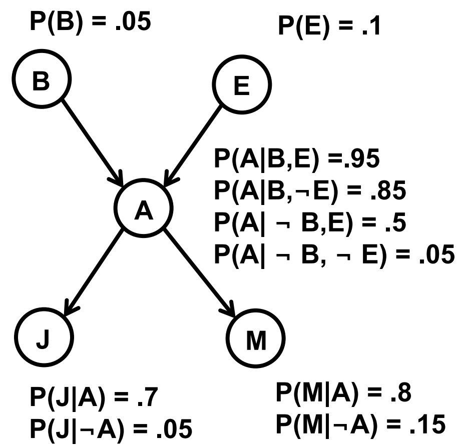
- We can reuse computations to reduce the running time
- However, there are still cases in which this algorithm will lead to exponential running time.
- Exact Bayesian Inference is NP-Hard.

Stochastic Inference

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

1. Sample the free variables
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



Using Sampling For Inference

- Revisiting our problem to compute P(B | J,¬M)
- Looking at the samples we can count:
  - 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:

$$P(B \mid J,\neg M) = P(B,J,\neg M) / P(J,\neg M) \approx N_B / N_c$$

Using Sampling For Inference

- Lets revisit our problem to compute P(B | J,¬M)
- Looking at the samples we can count:
  - *N*: total number of samples
  - N_c : total number of samples in w
  - N_B: total number of samples whe
- For a large enough N:
  - N_c / N ≈ P(J,¬M)
  - N_B / N ≈ P(B,J,¬M)
- And so, we can set:

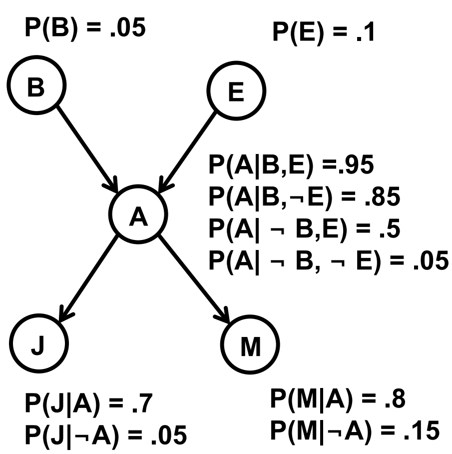
$$P(B \mid J,\neg M) = P(B,J,\neg M) / P(J,\neg M) \approx N_B / N_c$$

Problem: What if the condition rarely happens?

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

Weighted Sampling

- Again: P(B | J,¬M)
- Given an assignment to parents, we assign a value of 1 to J and 0 to M
- This way, all samples will contain the correct values for the conditional variables
- We record the *probability* of this assignment ($w = p_1 * p_2$) and we weight the new joint sample by *w*



Summary

- Bayes Rule
- Joint distribution, independence, conditional independence
- Attributes of Bayesian networks
- Constructing a Bayesian network
- Inference in Bayesian networks
  – exact, stochastic, weighted stochastic