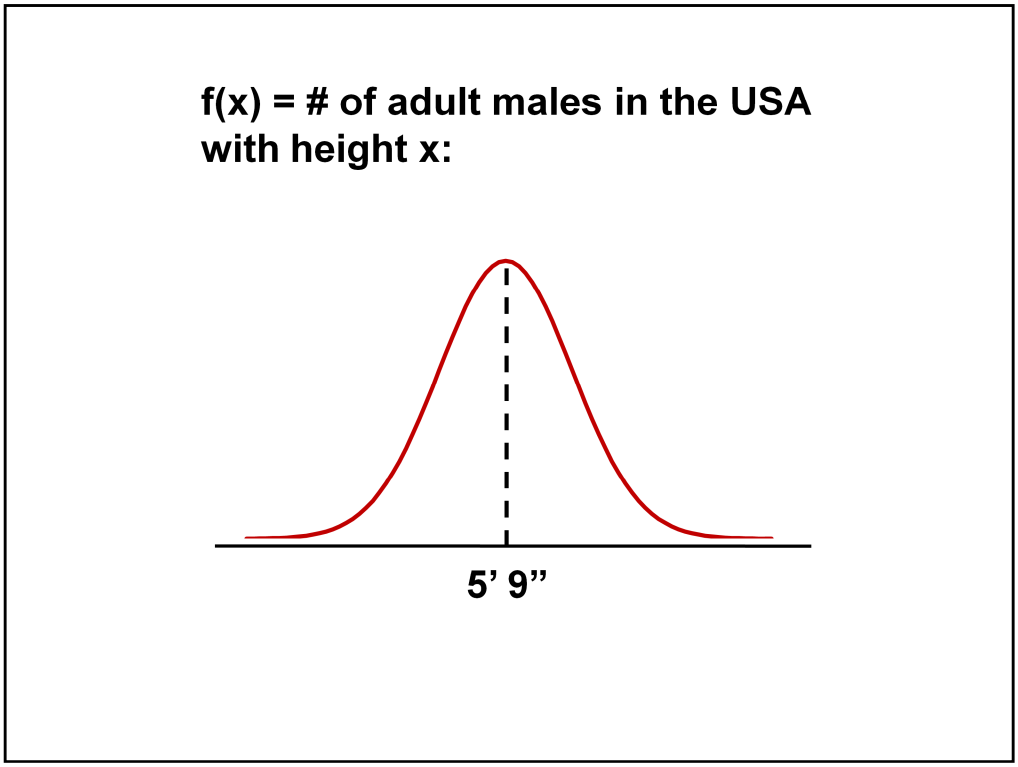
# Lecture 10: Bayesian Networks – Construction and Inference

(Russell and Norvig Chapter 14)




Probability distribution if area under curve = 1

Percentage of people shorter than 5’8’’: area under the curve for –inf to 5’8’’

# Gaussian (Normal) Distribution

**Center of the bell curve**

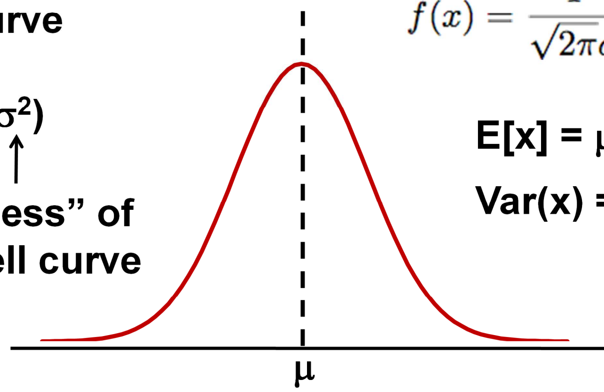
↓

**N($\mu$,$\sigma^2$)**

↑

**"Fatness" of the bell curve**

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{\frac{-(x-\mu)^2}{2\sigma^2}}$$




**E[x] = $\mu$**

**Var(x) = $\sigma^2$**

Two unbound variables in f(x) for a gaussian curve: sigma and mu

# Gaussian (Normal) Distribution

**A lot of stuff is normally distributed!**

**Central Limit Theorem (loosely): Sum of a large number of IID random variables is approximately Gaussian**

This is why a lot of stuff you measure looks like a Gaussian

# Independence

| Liked movie | Slept in movie | Raining | 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**

**In this case, the extra knowledge about rain does not change our prediction. Slept and rain are independent!**

The joint distribution table gives you all the information you need to ask these types of questions.

# Independence

**Events R and S are independent if: $P(S \mid R) = P(S)$**

**From this we can derive:**

- $P(\neg S \mid R) = P(\neg S)$
- $P(S,R) = P(S)P(R)$
- $P(R \mid S) = P(R)$

# 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**

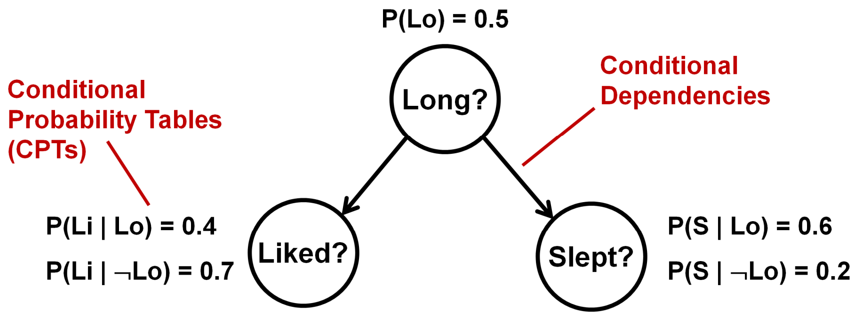
**P(slept, like movie | long) = 0.24**

**Given knowledge of length, the two other variables become independent**

A and B might be dependent on each other, however they are independent given c. Therefore they are conditionally independent

# Bayesian Networks

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




Bayesian networks allow you to specify the joint distribution without needing a huge table.

P(liked) = P(Li AND Lo) + P(Li AND not Lo) = P(Li|Lo)*P(Lo) + P(Li|not Lo)*P(not Lo) = 0.4*0.5 + 0.7*0.5 = 0.55

# 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 reconstruct the network for this problem?**

The scenario: My house is in LA equipped with an alarm, and I live in Pittsburgh. Mary & John are my neighbors in LA. If the alarm sounds, Mary or John might call me. Either a burglary or an earthquake can set off the alarm.

# Factoring Joint Distributions

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

**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)**

**This type of conditional dependencies can also be represented graphically**

First, this is a naïve way of calculating the joint distribution

# 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: $2^4$**

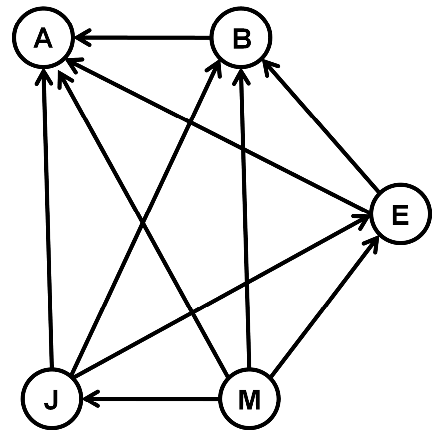
**B: $2^3$**

**E: 4**

**J: 2**

**M: 1**

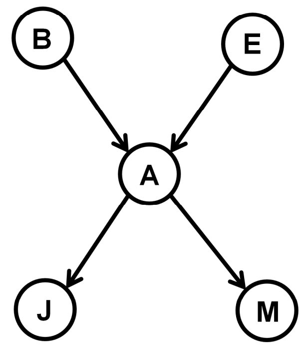
**Total: 31 parameters**



# 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**




**Number of parameters:**

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

**Total: 10 parameters**

**By relying on domain knowledge we saved 21 parameters!**

We are assuming here that Burglaries and earthquakes are independent (in the real world people might loot after an earthquake.)

## Constructing a Bayesian Network: Revisited

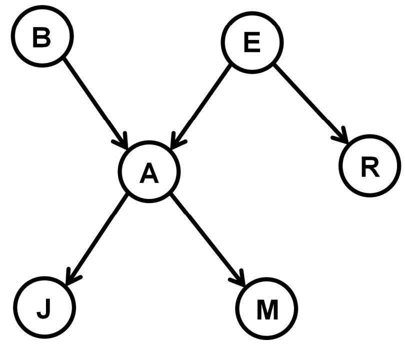
- Step 1: Identify the random variables
- Step 2: Determine the conditional dependencies
  - Select an ordering of the variables
  - Add them one at a time
  - For each new variable X added select the minimal subset of nodes as parents such that X is independent from all other nodes in the current network given its parents
- Step 3: Populate the CPTs

# Reconstructing Networks

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 algorithms 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?**

When you draw the bayes net, you are making assumptions about independence (like burglaries and earthquakes are independent). After it is drawn, you can ask any question.

# Inference

**Lets start with a simpler question: How can we compute a joint distribution from the network?**

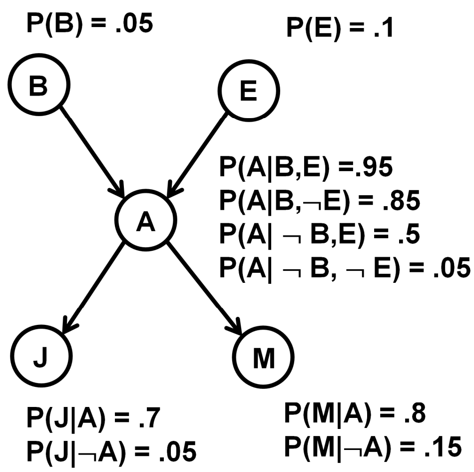
**For example, P(B,¬E,A,J, ¬M).**

**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**



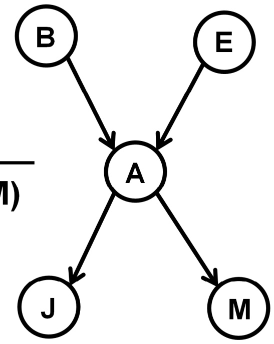
# Inference

We are now interested in queries of the form: $P(B \mid J, \neg M)$

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?



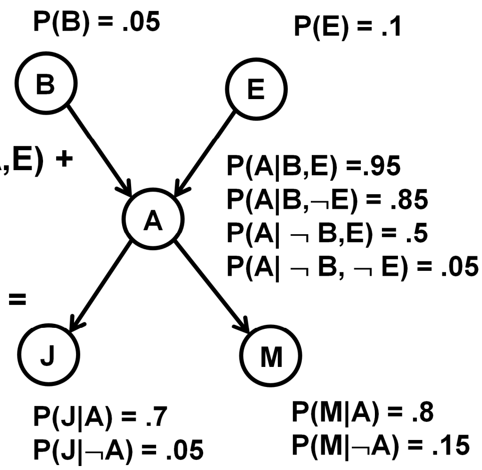
# Computing: $P(B,J,\neg M)$

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

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

$0.0007+0.00001+0.005+0.0003 =$

$0.00601$



# 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)**

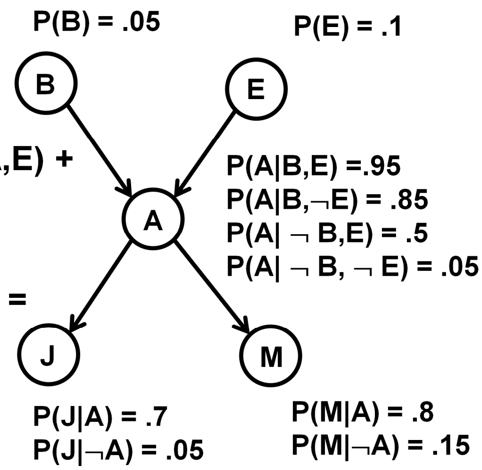
# Computing: P(B,J, ¬M)

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?**

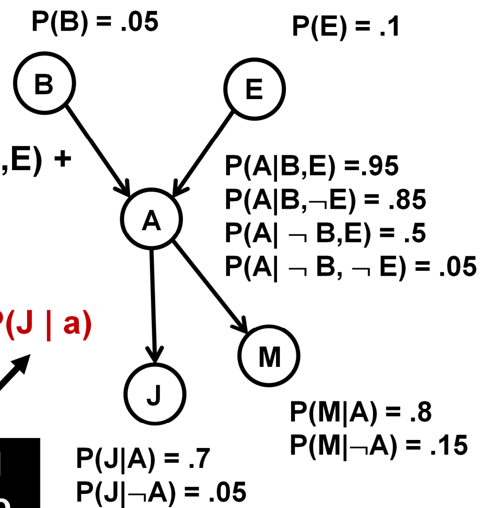
# Computing: P(B,J, ¬M)

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 because it includes inference in propositional logic as a special case**

However, on average using bayesian networks saves us time

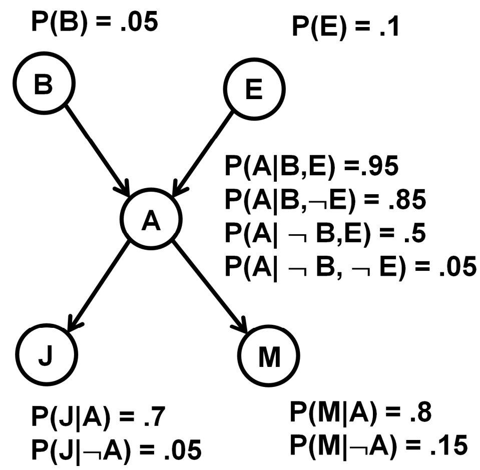
## Stochastic Inference

**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**




Prove that it must be possible

# 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 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 i**
  **(J,¬M)**

  **- $N_B$: total number of samples w**
- **For a large enough N:**

  **- $N_c$ / N ≈ P(J,¬M)**

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

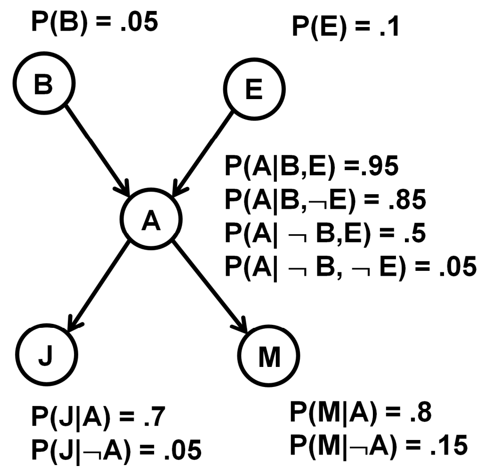
**P(B | J,¬M) = P(B,J,¬M) / P(J,¬M) ≈ $N_B$ / $N_c$**

**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)**
- **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***




P(J=1|…) = 0.05, P(M=0|…) = 0.85, so we're going to weigh this sample by 0.05*0.85

Luis strongly recommends you to read Ch.14 of Russell and Norvig

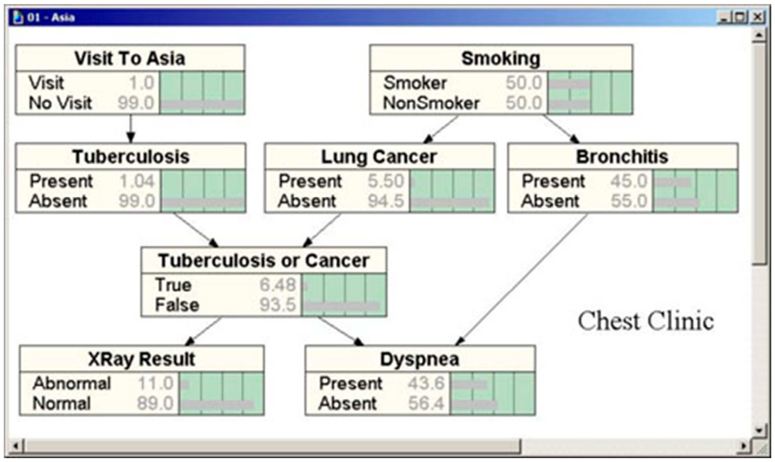
# Weighted Sampling Algorithm for Computing P(B | J,¬M)

- **Set $N_B$,$N_c$ = 0**
- **Repeat:**
  - **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:**

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

# 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

## Don’t Take so Many Classes!

**If you want to go to grad school, do research**

**If you want to make money, start a company**

**Ben Maurer**

Luis’s special public announcement