CSCI 5582
Artificial Intelligence

Lecture 14

Jim Martin

Today 10/17

- Review basics
- More on independence
- Break
- Bayesian Belief Nets

Review

- Joint Distributions
- Atomic Events
- Independence assumptions

Review: Joint Distribution

	Toothache=True	Toothache=False
Cavity True	0.04	0.06
Cavity False	0.01	0.89

•Each cell represents a conjunction of the variables in the model.

Atomic Events

- The entries in the table represent the probabilities of atomic events
  - Events where the values of all the variables are specified

Independence

- Two variables A and B are independent iff P(A|B) = P(A). In other words, knowing B gives you no information about B.
- Or P(A^B)=P(A|B)P(B)=P(A)P(B)
  - I.e. Two coin tosses

Mental Exercise

- With a fair coin which of the following two sequences is more likely?
  - HHHHHTTTTT
  - HTTHHHTHTT

Conditional Independence

- Consider the dentist problem with 3 variables: cavity, toothache, catch
- If I have a cavity, then the chances that there will be a catch is independent of whether or not I have a toothache as well. I.e.
  - P(Catch|Cavity^Toothache)= P(Catch|Cavity)

Conditional Independence

- Remember that having the joint distribution over N variables allows you to answer all the questions involving those variables.
- Exploiting conditional independence allows us to represent the complete joint distribution with fewer entries.
  - I.e. Fewer than the 2^N normally needed

Conditional Independence

- P(Cavity,Catch,Toothache)

= P(Cavity)P(Catch,Toothache|Cavity)

=P(Cavity)P(Catch|Cavity)P(Toothache|Cavity)

Conditional Independence

- P(Cavity,Catch,Toothache)

= P(Catch)P(Cavity,Toothache|Catch)

⇒Huh?

Bayesian Belief Nets

- A compact notation for representing conditional independence assumptions and hence a compact way of representing a joint distribution.
- Syntax:
  - A directed acyclic graph, one node per variable
  - Each node augmented with local conditional probability tables

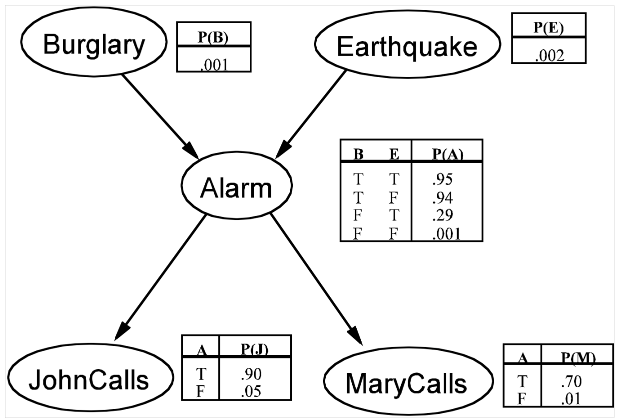
Bayesian Belief Nets

- Nodes with no incoming arcs (root nodes) simply have priors associated with them
- Nodes with incoming arcs have tables enumerating the
  - P(Node|Conjunction of Parents)
  - Where parent means the node at the other end of the incoming arc

Alarm Example

- Variables: *Burglar, MaryCalls, JohnCalls, Earthquake, Alarm*
- Network topology captures the domain causality (conditional independence assumptions).

Alarm Example



Bayesian Belief Nets: Semantics

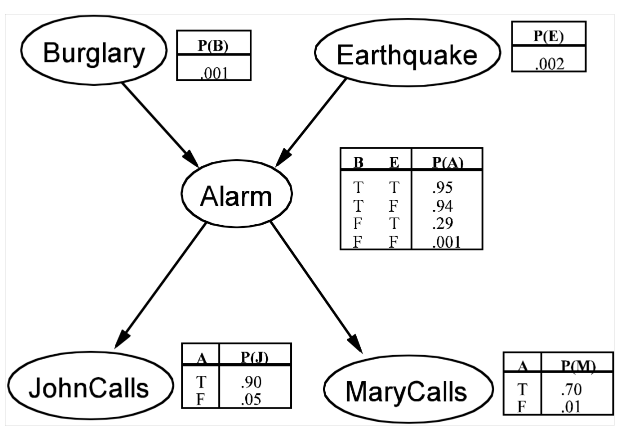
- The full joint distribution for the N variables in a Belief Net can be recovered from the information in the tables.

$$P(X_1,\ldots X_N) = \prod_{i=1}^{N} P(X_i \mid Parents(X_i))$$

Belief Net Semantics
Alarm Example

- What are the chances of John calls, Mary calls, alarm is going off, no burglary, no earthquake?

Alarm Example



Alarm Example

- P(J^M^A^~B^~E)=

  P(J|A)*P(M|A)*P(A|~B^~E)*P(~B)*P(~E)
  0.9 * 0.7 * .001 * .999 * .998

- In other words, the probability of atomic events can be read right off the network as the product of the probability of the entries for each variable

Events

- What about non-atomic events?
- Remember to partition. Any event can be defined as a combination of other more well-specified events.

  P(A) = P(A^B)+P(A^~B)
- So what's the probability that Mary calls out of the blue?

Events

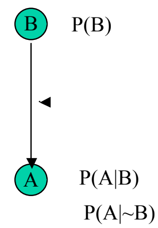
- P(M ^J^E^B^A)+
  P(M^J^E^B^~A)+
  P(M^J^E^~B^A)+
  …

Events

- How about P(M|Alarm)?
  - Trick question… that's something we know
- How about P(M|Earthquake)?
  - Not directly in the network
    rewrite as
    P(M^Earthquake)/P(Earthquake)

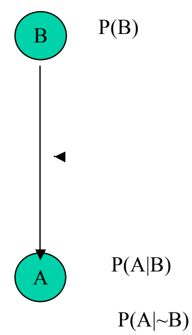
Simpler Examples

- Let's say we have two variables A and B, and we know B influences A.
- What's P(A^B)?



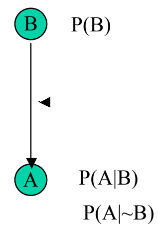
Simple Example

- Now I tell you that B has happened.
- What's you belief in A?



Simple Example

- Suppose instead I say A has happened
- What's you belief in B?



Simple Example

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

  = P(B^A)/P(A^B)+P(A^~B)

$$= \frac{P(B)P(A|B)}{P(B)P(A|B)+P(\sim B)P(A|\sim B)}$$

Chain Rule Basis

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

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

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

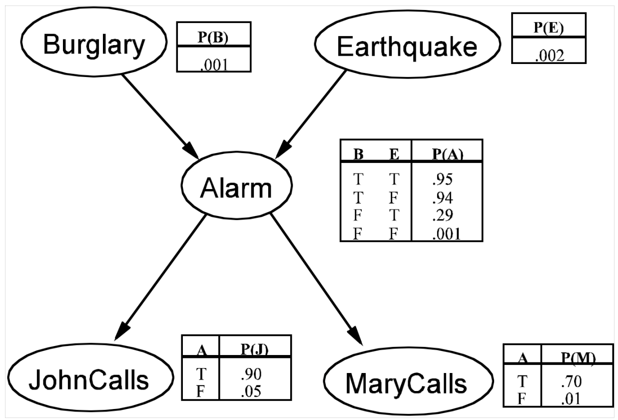
P(A|B,E)P(B,E)

P(B|E)P(E)

Chain Rule Basis

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

Alarm Example



Details

- Where do the graphs come from?
  - Initially, the intuitions of domain experts
- Where do the numbers come from?
  - Hopefully, from hard data
  - Sometimes from experts intuitions
- How can we compute things efficiently?
  - Exactly by not redoing things unnecessarily
  - By approximating things

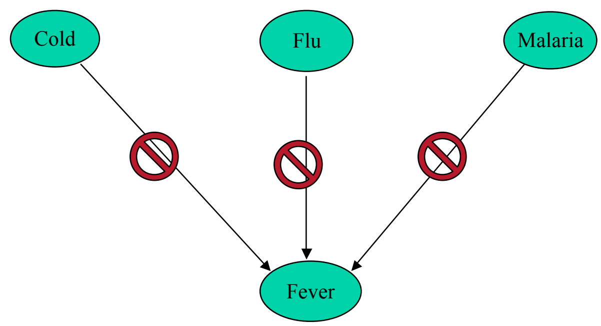
Break

- Readings for probability
  - 13: All
  - 14:
    - 492-498, 500, Sec 14.4

Noisy-Or

- Even with the reduction in the number of probabilities needed it's hard to accumulate all the numbers you need.
- Especially true when some evidence variables are shared among many causes.
- The Noisy-Or hack is a useful short-cut.
- P(A|C1^C2^C3)

Noisy-Or



Noisy Or

- P(Fever|Cold)
- P(Fever|Malaria)
- P(Fever|Flu)

- P(~Fever|Cold)
- P(~Fever|Malaria)
- P(~Fever|Flu)

Noisy Or

- What does it mean for the  to occur?
- It means the cause was true and the symptom didn't happen
- What's the probability of that?
  - P(~Fever|Cause)
    - P(~Fever|Flu), etc

Noisy Or

- If all three causes are true and you don't have a fever then all three blockers  are in effect
- What's the probability of that?
  - P(~Fever|flu,cold,malaria)
  - P(~Fever|flu)P(~Fever|cold)P(~Fever|malaria)
- But 1 – that = P(Fever|causes)

Computing with BBNs

- Normal scenario
  - You have a belief net consisting of a bunch of variables
    - Some of which you know to be true (evidence)
    - Some of which you're asking about (query)
    - Some you haven't specified (hidden)

Example

- Probability that there's a burglary given that John and Mary are calling
- P(B|J,M)
  - B is the query variable
  - J and M are evidence variables
  - A and E are hidden variables

Example

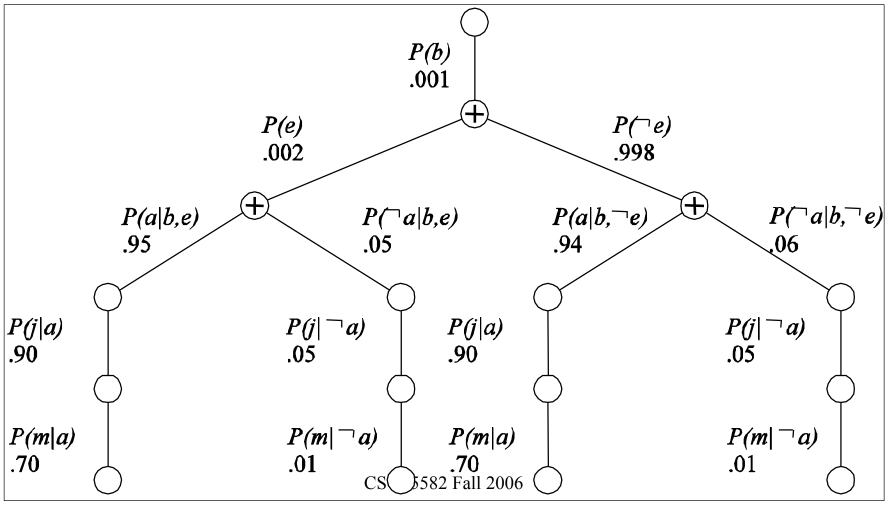
- Probability that there's a burglary given that John and Mary are calling
- P(B|J,M) = alpha P(B,J,M)
  = alpha *
  P(B,J,M,A,E) +
  P(B,J,M,~A,E)+
  P(B,J,M,A,~E)+
  P(B,J,M, ~A,~E)

From the Network

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

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

Expression Tree



Speedups

- Don't recompute things.
  - Dynamic programming
- Don't compute some things at all
  - Ignore variables that can't effect the outcome.

Example

- John calls given burglary
- P(J|B)

$$\alpha P(B)\sum_e P(E)\sum_a P(A \mid B,E)P(J \mid a)\sum_m P(M \mid A)$$

Variable Elimination

- Every variable that is not an ancestor of a query variable or an evidence variable is irrelevant to the query

Next Time

- Finish Chapters 13 and 14