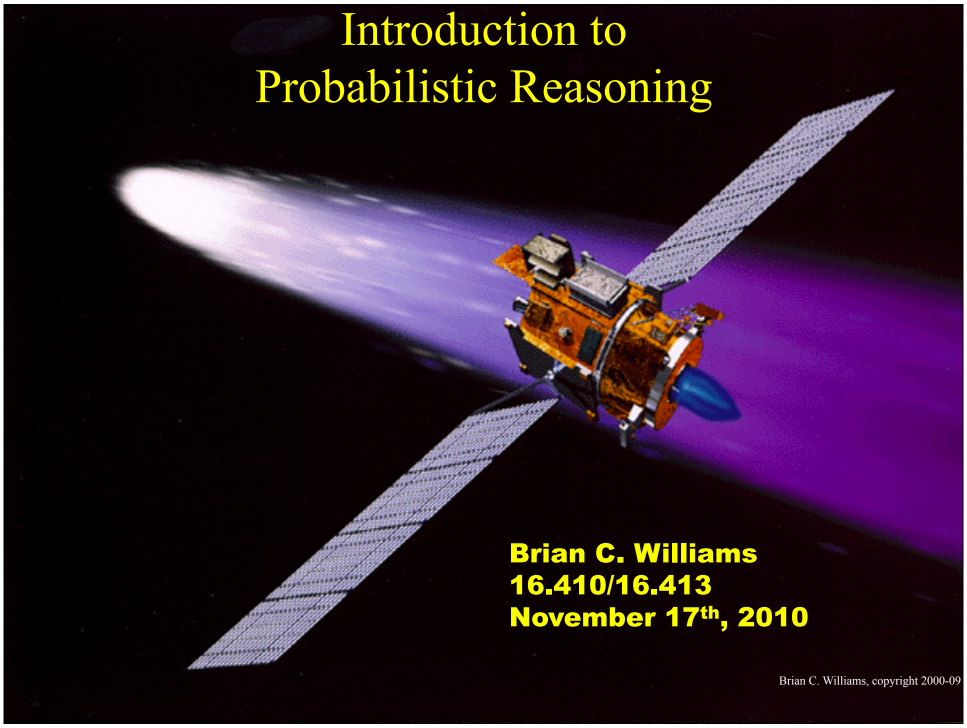


Image credit: NASA.

Assignment

- Homework:
  - Problem Set #8: Linear Programming, due today, Wednesday, November 16th.
  - Problem Set #9: Probabilistic Reasoning, out today, due Wednesday, November 24th.
- Readings:
  - Today: Review of Probabilities and Probabilistic Reasoning.
    - AIMA Chapter 13.
    - AIMA Chapter 14, Sections. 1-5.
  - Monday: HMMs, localization & mapping
    - AIMA Chapter 15, Sections. 1-3.

11/17/10 2

Notation

- S, Q, R, P — Logical sentences
- Φ — Background theory (a sentence).
- not, and (∧), or (v), implies (→), "if and only if " (iff, ≡). Standard logical connectives where iff ≡ "if and only if".
- M(S), entails, ⊥ — Models of sentence S, entails, false.

- A, B, C — Sets.
- U, ϕ — Universe of all elements, empty set.

- ∪, ∩, ~, - — Set union, intersection, inverse and difference.
- ≡ — Equivalent to.

- V: — Variable or vector of variables.
- V_i: — The ith variable of vector V.
- V, v_i: — A particular assignment to V; short form for V= v, $V_i = v_i$.
- V^t: — V at time t.
- $V^{i:j}$: — A sequence of V from time i to time j.

Notation

- S: — States or state vector.
- O: — Observables or observation vector.
- X: — Mode or mode vector.
- S0, S1 — Stuck at 0/1 mode.
- Prefix (k) L — Returns the first k elements of list L.
- Sort L by R — Sorts list L in increasing order based on relation R.

- s_i — ith sample in sample space U.
- P(X) — The probability of X occurring.
- P(X|Y) — The probability of X, conditioned on Y occurring.
- A ⊥ C | B) — A is *conditionally independent* of C given B.

Outline

- Motivation
- Set Theoretic View of Propositional Logic
- From Propositional Logic to Probabilities
- Probabilistic Inference
  - General Queries and Inference Methods
  - Bayes Net Inference
  - Model-based Diagnosis (Optional)

Multiple Faults Occur

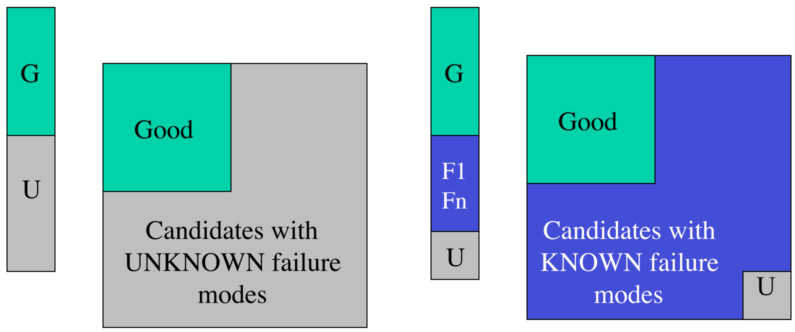
Image source: NASA. APOLLO 13

- three shorts, tank-line and pressure jacket burst, panel flies off.

Lecture 16: Framed as CSP.

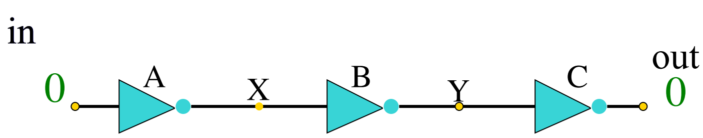
- How do we compare the space of alternative diagnoses?
- How do we explain the cause of failure?
- How do we prefer diagnoses that explain failure?

Due to the unknown mode, there tends to be an exponential number of diagnoses.



1. Introduce fault models.
   - More constraining, hence more easy to rule out.
   - Increases size of candidate space.
2. Enumerate most likely diagnoses X_i based on probability.
   Prefix (k) (Sort $\{X_i\}$ by decreasing $P(X_i \mid O)$)

- Most of the probability mass is covered by a few diagnoses.



Idea: Include known fault modes (S0 = “stuck at 0,” S1=“stuck at 1,”) as well as Unknown.

Diagnoses: (42 of 64 candidates)

Fully Explained Failures:
- [A=G, B=G, C=S0]
- [A=G, B=S1, C=S0]
- [A=S0, B=G, C=G]

. . .

Partially Explained:
- [A=G, B=U, C=S0]
- [A=U, B=S1, C=G]
- [A=S0, B=U, C=G]

. . .

Faults Isolated, No Explanation:
- [A=G, B=G, C=U]
- [A=G, B=U, C=G]
- [A=U, B=G, C=G]

Estimate Dynamically with a Bayes Filter

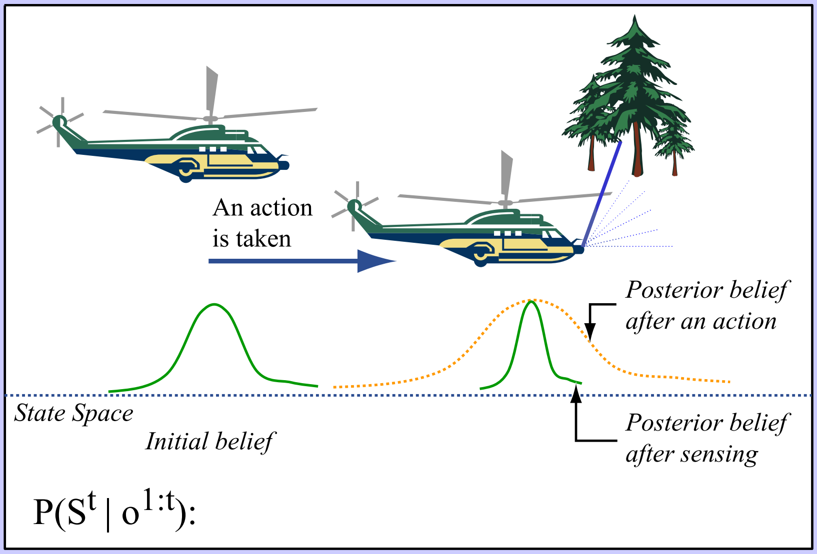
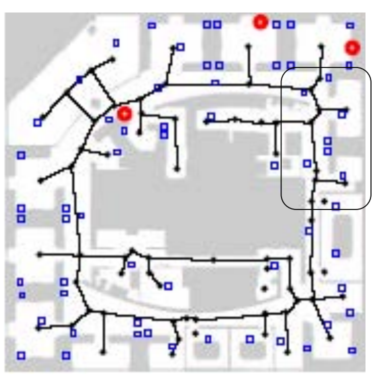
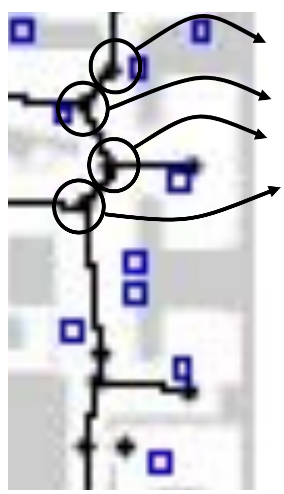


Image by MIT OpenCourseWare.

Localizing a Robot within a Topological Map





x_1

x_2: $p(x_2|x_1,a)= .9$

x_3: $p(x_3|x_1,a)=.05$

x_4: $p(x_4|x_1,a)=.05$

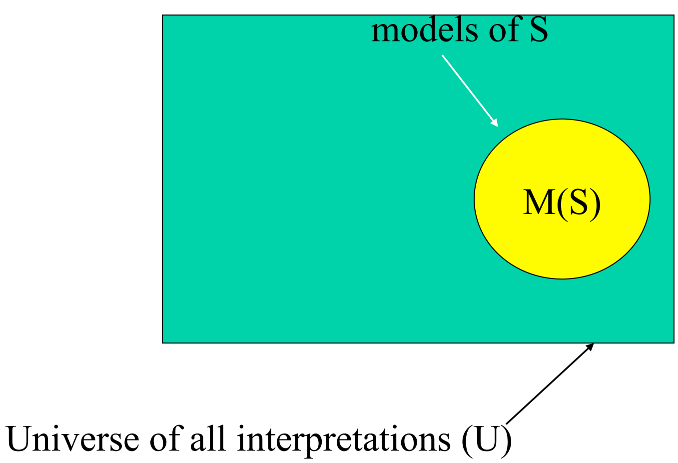
Observations can be features such as corridor features, junction features, etc.

Outline

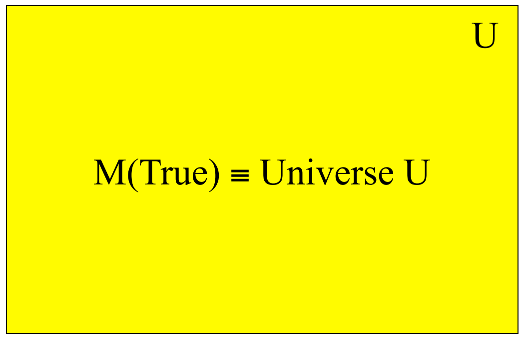
- Motivation
- Set Theoretic View of Propositional Logic
- From Propositional Logic to Probabilities
- Probabilistic Inference

Propositional Logic
Set Theoretic Semantics

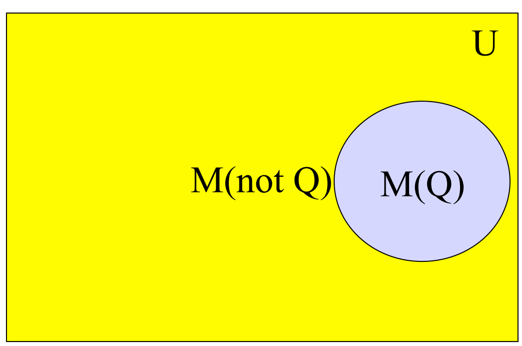
Given sentence S:



Set Theoretic Semantics: S ≡ True

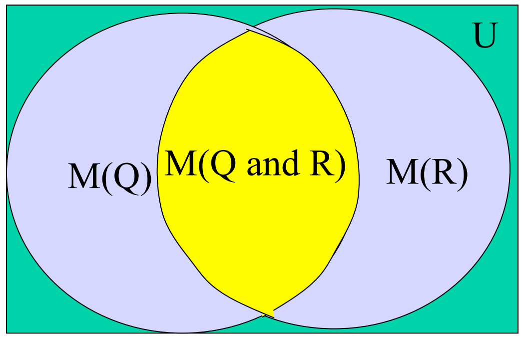


Set Theoretic Semantics: S ≡ not Q



M(not Q) ≡ U – M(Q)

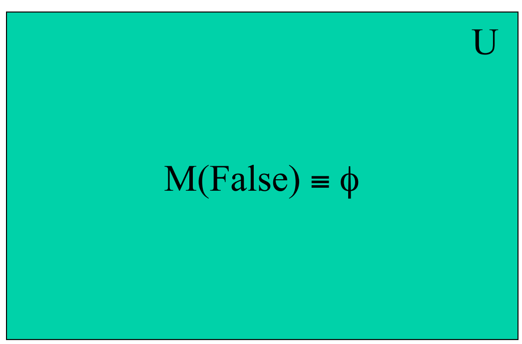
Set Theoretic Semantics: S ≡ Q and R



M(Q and R) ≡ M(Q) ∩ M(R)

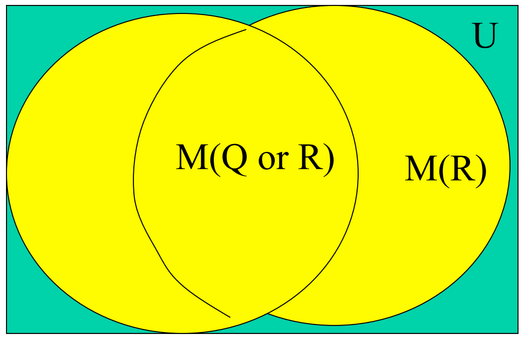
11/17/10 15

Set Theoretic Semantics: False



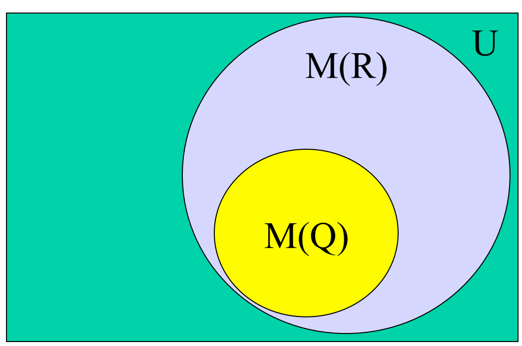
11/17/10 16

Set Theoretic Semantics: Q or R



M(Q or R) ≡ M(Q) ∪ M(R)

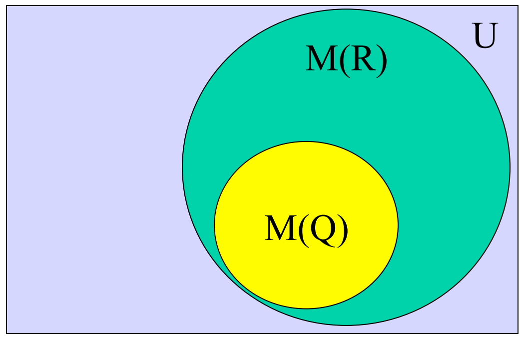
Set Theoretic Semantics: Q implies R



"Q implies R" is True iff Q entails R

M(Q implies R) ≡ M(Q) ⊆ M(R)

Set Theoretic Semantics: P implies Q



"Q implies R" is True iff "Q and not R" is inconsistent

"Q implies R" is True iff M(Q and not R) $\equiv \phi$

11/17/10 19

Axioms of Sets over Universe U

1.	$A \cup B \equiv B \cup A$	Commutativity
2.	$A \cup (B \cup C) \equiv A \cup (B \cup C)$	Associativity
3.	$A \cap (B \cup C) \equiv A \cap B \cup A \cap C$	Distributivity
4.	$\sim(\sim A) \equiv A$	~ Elimination
5.	$\sim(A \cap B) \equiv (\sim A) \cup (\sim B)$	De Morgan's
6.	$A \cap (\sim A) \equiv \phi$	
7.	$A \cap U \equiv A$	Identity

$\Rightarrow$ propositional logic axioms follow from these axioms.

11/17/10 20

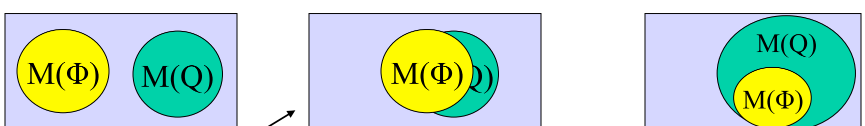
Outline

- Motivation
- Set Theoretic View of Propositional Logic
- From Propositional Logic to Probabilities
  - Degrees of Belief
  - Discrete Random Variables
- Probabilistic Inference

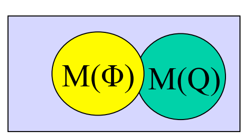
Why Do We Need Degrees of Belief?

"Given theory Φ, sentence Q …."

"… is never true." "… could be true." "… must be true."



Q = "One of the valves to Engine A is stuck closed."



Q = "All of the components in the propulsion system are broken in a way never seen before."

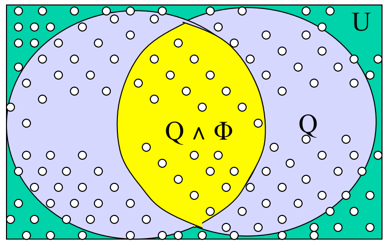
unsatisfiable satisfiable entailment

Degrees of Belief

Probability P: Events → [0,1]
is a measure of the likelihood that an Event is true.

Events:
- like propositions.

Φ, Q, Φ ∧ Q



Sample Space U ≡ $\{s_i\}$:
- like interpretations.
- mutually exclusive,
- collectively exhaustive,
- finest grained events.
- P defined over $s_i \in U$.

$$P(\Phi) = \sum_{s_i \in \Phi} P(s_i) \qquad P(U) = \sum_{s_i \in U} P(s_i) = 1 \qquad P(Q) = \sum_{s_i \in Q} P(s_i)$$

$$P(\Phi \wedge Q) = \sum_{s_i \in \Phi \cap Q} P(s_i)$$

Like counting *weighted models* in logic.

11/17/10 23

Axioms of Probability

P: Events → Reals

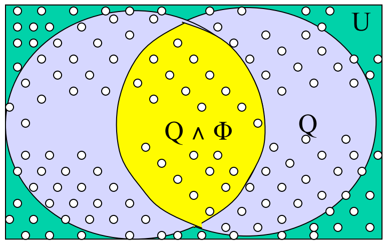
1. For any event A, P(A) ≥ 0.
2. P(U) = 1.
3. If A ∩ B = ϕ, then P(A ∪ B) = P(A) + P(B).

⇨ All conventional probability theory follows.

11/17/10 24

Conditional Probability: P(Q | Φ)

Is Q true given Φ? ⇨ What is P("Q given Φ")?



$P(s_j \mid \Phi) = P(s_j) / P(\Phi)$ if $s_j \in \Phi$

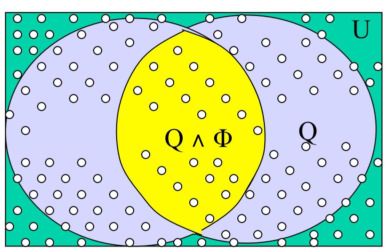
0 otherwise

$= \dfrac{P(s_j, \Phi)}{P(\Phi)}$

11/17/10 25

Conditional Probability: P(Q | Φ)

Is Q true given Φ? ⇨ What is P("Q given Φ")?



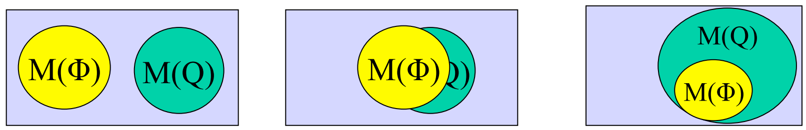
$$P(Q \mid \Phi) = \sum_{s_j \in Q} P(s_j \mid \Phi) = \sum_{s_j \in Q} \frac{P(s_j, \Phi)}{P(\Phi)} = \frac{P(Q, \Phi)}{P(\Phi)}$$

11/17/10 26

Degree of Belief

"Given theory Φ, sentence Q …."

"… is inconsistent." "… could be true." "… must be true."



$P(Q \mid \Phi)=$ 0 $\frac{P(Q, \Phi)}{P(\Phi)}$ 1

Representing Sample Space Using Discrete Random Variables

Discrete Random Variable X :

- Domain $D_X = \{x_1, x_2, \dots x_n\}$
  - mutually exclusive, collectively exhaustive.
- P(X): $D_X \rightarrow [0, 1]$
  - $\sum P(x_i) = 1$

Joint Distribution over $X_1, \dots X_n$:

- Domain $\prod D_{Xi}$
- $P(X_1, \dots X_n)$: $\prod D_{Xi} \rightarrow [0, 1]$
  - Notation: $P(x_1, \dots x_n) \equiv P(X_1 = x_1 \wedge \dots . \wedge X_n = x_n)$

Y ∈ {cloudy,	.197	.20
sunny}	.003	.60

Z ∈ {raining, dry}

Outline

- Motivation
- Set Theoretic View of Propositional Logic
- From Propositional Logic to Probabilities
- Probabilistic Inference
  - General Queries and Inference Methods
  - Bayes Net Inference
  - Model-based Diagnosis (Optional)

Types of Probabilistic Queries

Let X = <S; O>

- Belief Assessment
  - $b(S_i) = P(S_i \mid o)$
- Most Probable Explanation (MPE)
  - $s^* = \arg\max P(s \mid o)$ for all $s \in D_S$
- Maximum Aposteriori Hypothesis (MAP)
  - Given $A \subseteq S$
    $a^* = \arg\max P(a \mid o)$ for all $a \in D_A$
- Maximum Expected Utility (MEU)
  - Given decision variables D
    $d^* = \arg\max \sum_x u(x)P(x \mid d)$ for all $d \in D_D$

Common Inference Methods and Approximations used to Answer Queries

1. Product (chain) rule.
2. Product rule + conditional independence.
3. Eliminate variables by marginalizing.
4. Exploiting distribution during elimination.
5. Conditional probability via elimination.
6. Conditional probability via Bayes Rule.

Approximations:

1. Uniform Likelihood Assumption
2. Naïve Bayes Assumption

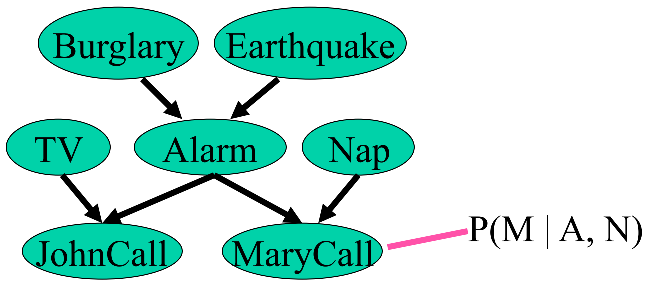
11/17/10 31

Outline

- Motivation
- Set Theoretic View of Propositional Logic
- From Propositional Logic to Probabilities
- Probabilistic Inference
  - General Queries and Inference Methods
  - Bayes Net Inference
  - Model-based Diagnosis (Optional)

11/17/10 32

Bayesian Network

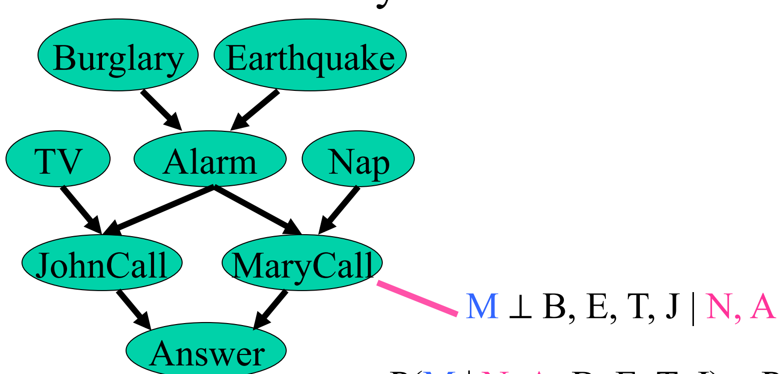


Input: Directed acyclic graph:

- Nodes — denote random variables.
- Arcs — tails denote parents P of child C.
- Conditional probability — P(C | P) for each node.

Output: Answers to queries given earlier.

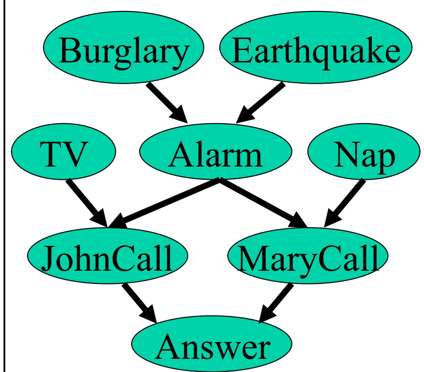
Conditional Independence for a Bayesian Network



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

A variable is *conditionally independent* of its *non-descendants*, given its *parents*.

Computing Joint Distributions P(X) for Bayesian Networks



Joint Distribution:

- Assigns probabilities to every interpretation.
- Queries computed from joint.

How do we compute joint distribution P(J,M,T,A,N,B,E)?

Product (Chain) Rule (Method 1)

$$P(C,S) = P(C \mid S)P(S)$$

C (Cancer) ∈ {none, benig, malig}
S (Smoking) ∈{no, light, heavy}

P(C| S) :

Smoking =	no	light	heavy
P(C=none)	0.96	0.88	0.60
P(C=benign)	0.03	0.08	0.30
P(C=malig)	0.01	0.04	0.10

P(S) :

P(S=no)	0.80
P(S=light)	0.15
P(S=heavy)	0.05

= 0.96 x 0.80

P(C, S):

Smoking / Cancer	none	benign	malig
no	0.768	0.024	0.008
light	0.132	0.012	0.006
heavy	0.03	0.015	0.005

Product Rule + Conditional Independence (Method 2)

If A is *conditionally independent* of C given B (i.e., $A \perp C \mid B$), then:

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

Suppose $A \perp B, C \mid E$ and $B \perp C \mid E$, find $P(A, B, C \mid E)$:

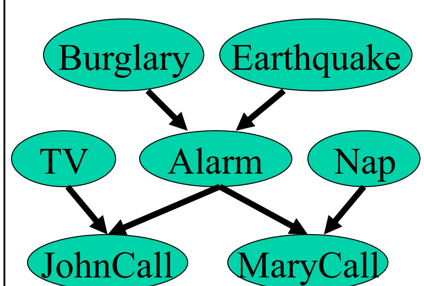
$$P(A, B, C \mid E) = P(A \mid \cancel{B}, \cancel{C}, E) P(B \mid \cancel{C}, E) P(C \mid E)$$

Product rule

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

Independence

Product Rule for Bayes Nets



- X_i is independent of its ancestors, given its parents:

$$P(X) = \prod_{X_i \in X} P(X_i \mid Parents(X_i))$$

Product rule

$$P(J, M, T, A, N, B, E)$$
$$= P(J \mid \cancel{M}, T, A, \cancel{N}, \cancel{B}, \cancel{E}) P(M \mid \cancel{T}, A, N, \cancel{B}, \cancel{E}) \ldots P(B \mid E) P(E)$$
$$= P(J \mid T, A) P(M \mid A, N) P(A \mid B, E) P(T) P(N) P(B) P(E)$$

Independence

Computing Probabilistic Queries

Let X = <S; O>

Belief Assessment:

– $b(S_i) = P(S_i \mid e)$

Most Probable Explanation (MPE):

– $s^* = \arg\max P(s \mid o)$ for all $s \in D_S$

Maximum Aposteriori Hypothesis (MAP):

– Given $A \subseteq S$
  $a^* = \arg\max P(s \mid o)$ for all $a \in D_A$

Solution: Some combination of

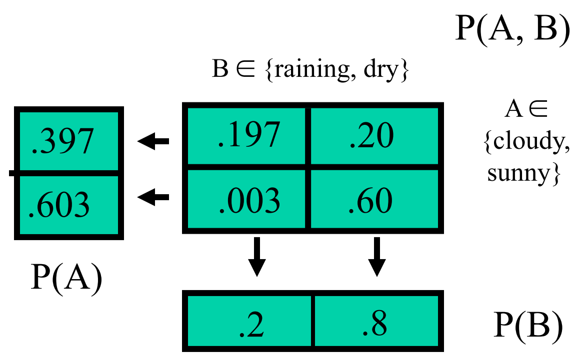
1. Start with joint distribution.
2. Eliminate some variables *(marginalize)*.
3. Condition on observations.
4. Find assignment maximizing probability.

11/17/10 copyright Brian Williams, 2005-10 39

Eliminate Variables by Marginalizing (Method 3)

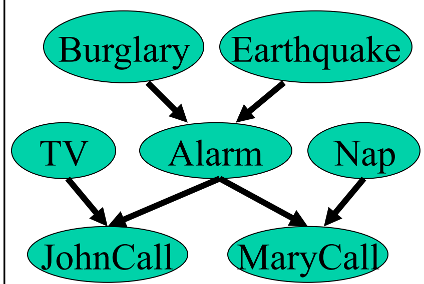
Given P(A, B)
find P(A) :

$$P(A) = \sum_{b_i \in B} P(A, b_i)$$



11/17/10 copyright Brian Williams, 2005-10 40

Exploit Distribution During Elimination (Method 4)



- Bayes Net Chain Rule: X_i is independent of its ancestors, given its parents.

$$P(X) = \prod_{X_i \in X} P(X_i \mid Parents(X_i))$$

Issue: Size of P(X) is exponential in |X|
Soln: Move sums inside product.

$$P(J,M,B,E) = \sum_{A,T,N} P(J \mid T,A)P(M \mid A,N)P(A \mid B,E)P(N)P(B)P(E)P(T)$$

$$= P(B)P(E)\sum_{A} P(A \mid B,E)\sum_{T} P(J \mid T,A)P(T)\sum_{N} P(M \mid A,N)P(N)$$

Compute Conditional Probability via Elimination (Method 5)

$$P(A \mid B) = \frac{P(A,B)}{P(B)} = \frac{P(A,B)}{\sum P(B,a_i)} = \alpha P(A,B)$$

Note: The denominator is constant for all a_i in A

Example: P(A | B)?

P(A,B), with $B \in$ {raining, dry} across the columns and $B \in$ {cloudy, sunny} down the rows:

.197	.20
.003	.60

P(B):

.2	.8

P(A | B):

.197/.2	.20/.8
.003/.2	.60/.8

Computing Probabilistic Queries

Let X = <S, O>

Belief Assessment: $S = \langle S_i, Y \rangle$

– $b(S_i) = P(S_i \mid o)$

$$= \frac{\sum_{y \in D_Y} P(S_i, y, o)}{\sum_{s_i \in D_{S_i}, y \in D_Y} P(s_i, y, o)}$$

Most Probable Explanation (MPE):

– $s^* = \arg\max P(s \mid o)$ for all $s \in D_S$

$$= \arg\max_{s \in D_S} P(s, o)$$

Maximum Aposteriori Hypothesis (MAP): $S = \langle A, Y \rangle$

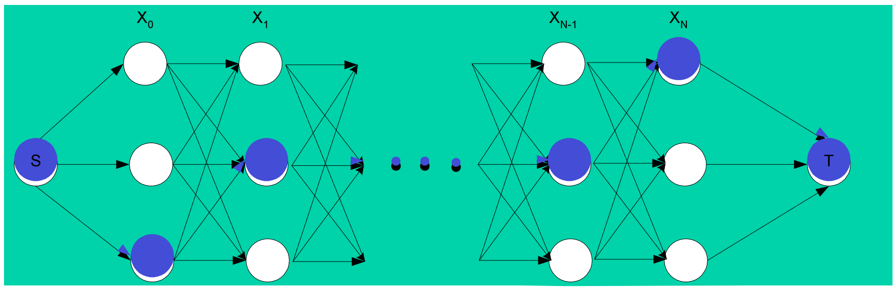
– Given $A \subseteq S$
  $a^* = \arg\max P(a \mid o)$ for all $a \in D_A$

$$= \arg\max_{a \in D_A} \sum_{y \in D_Y} P(a, y, o)$$

Outline

- Motivation
- Set Theoretic View of Propositional Logic
- From Propositional Logic to Probabilities
- Probabilistic Inference
  - General Queries and Inference Methods
  - Bayes Net Inference
  - Model-based Diagnosis (Optional)

Estimating States of Dynamic Systems (next week)



Given sequence of commands and observations:

- Infer current distribution of (most likely) states.
- Infer most likely trajectories.