# Reasoning Under Uncertainty: Conditioning, Bayes Rule & Chain Rule

CPSC 322 – Uncertainty 2

Textbook §6.1.3

March 18, 2011

# Lecture Overview

- Recap: Probability & Possible World Semantics
- Reasoning Under Uncertainty
  - Conditioning
  - Inference by Enumeration
  - Bayes Rule
  - Chain Rule

# Course Overview

**Course Module:** *Representation* — Reasoning Technique

## Environment

| Problem Type | | Deterministic | Stochastic | |
|---|---|---|---|---|
| Static | Constraint Satisfaction | ***Variables + Constraints*** (Arc Consistency; Search) | | |
| Static | Logic | ***Logics*** (Search) | ***Bayesian Networks*** (Variable Elimination) | Uncertainty |
| Sequential | Planning | ***STRIPS*** (Search; As CSP (using arc consistency)) | ***Decision Networks*** (Variable Elimination); ***Markov Processes*** (Value Iteration) | Decision Theory |

For the rest of the course, we will consider uncertainty

# Recap: Possible Worlds Semantics

- Example: model with 2 random variables
  - Temperature, with domain {hot, mild, cold}
  - Weather, with domain {sunny, cloudy}
- One joint random variable
  - <Temperature, Weather>
  - With the crossproduct domain {hot, mild, cold} × {sunny, cloudy}
- There are 6 possible worlds
  - The joint random variable has a probability for each possible world

| *Weather* | *Temperature* | *µ(w)* |
|---|---|---|
| sunny | hot | 0.10 |
| sunny | mild | 0.20 |
| sunny | cold | 0.10 |
| cloudy | hot | 0.05 |
| cloudy | mild | 0.35 |
| cloudy | cold | 0.20 |

- We can read the probability for each subset of variables from the joint probability distribution
  - E.g. P(Temperature=hot) = P(Temperature=hot,Weather=Sunny)
    + P(Temperature=hot, Weather=cloudy)
    = 0.10 + 0.05

# Recap: Possible Worlds Semantics

- *Examples for* “⊧” (related but not identical to its meaning in logic)
  - $w_1$ ⊧ W=sunny
  - $w_1$ ⊧ T=hot
  - $w_1$ ⊧ W=sunny ∧ T=hot

- E.g. f = “T=hot”
  - *Only* $w_1$ ⊧ f and $w_4$ ⊧ f
  - p(f) = μ($w_1$) + μ($w_4$)
    = 0.10 + 0.05

- E.g. f ’ = “W=sunny ∧ T=hot”
  - *Only* $w_1$ ⊧ f ’
  - p(f ’) = μ($w_1$) = 0.10

| *Name of possible world* | *Weather W* | *Temperature T* | *Measure μ of possible world* |
|---|---|---|---|
| $w_1$ | sunny | hot | 0.10 |
| $w_2$ | sunny | mild | 0.20 |
| $w_3$ | sunny | cold | 0.10 |
| $w_4$ | cloudy | hot | 0.05 |
| $w_5$ | cloudy | mild | 0.35 |
| $w_6$ | cloudy | cold | 0.20 |

w ⊧ X=x means variable X is assigned value x in world w

- Probability measure μ(w) sums to 1 over all possible worlds w
- The probability of proposition f is defined by: $$p(f) = \sum_{w \models f} \mu(w)$$

# Recap: Probability Distributions

**Definition (probability distribution)**

A probability distribution P on a random variable X is a function dom(X) $\rightarrow$ [0,1] such that

$$x \rightarrow P(X=x)$$

Note: We use notations P(f) and p(f) interchangeably

# Recap: Marginalization

- Given the joint distribution, we can compute distributions over smaller sets of variables through marginalization:

$$P(X{=}x) = \Sigma_{z \in dom(Z)} P(X{=}x, Z = z)$$

- This corresponds to summing out a dimension in the table.
- The new table still sums to 1. It must, since it's a probability distribution!

| *Weather* | *Temperature* | *µ(w)* |
|---|---|---|
| sunny | hot | 0.10 |
| sunny | mild | 0.20 |
| sunny | cold | 0.10 |
| cloudy | hot | 0.05 |
| cloudy | mild | 0.35 |
| cloudy | cold | 0.20 |

| *Temperature* | *µ(w)* |
|---|---|
| hot | 0.15 |
| mild | |
| cold | |

P(Temperature=hot) =
  P(Temperature = hot, Weather=sunny)
+ P(Temperature = hot, Weather=cloudy)
= 0.10 + 0.05 = 0.15

# Recap: Marginalization

- Given the joint distribution, we can compute distributions over smaller sets of variables through marginalization:

$$P(X=x) = \Sigma_{z \in dom(Z)} P(X=x, Z = z)$$

- This corresponds to summing out a dimension in the table.
- The new table still sums to 1. It must, since it's a probability distribution!

| *Weather* | *Temperature* | *µ(w)* |
|---|---|---|
| sunny | hot | 0.10 |
| sunny | mild | 0.20 |
| sunny | cold | 0.10 |
| cloudy | hot | 0.05 |
| cloudy | mild | 0.35 |
| cloudy | cold | 0.20 |

| *Temperature* | *µ(w)* |
|---|---|
| hot | 0.15 |
| mild | 0.55 |
| cold | |

# Recap: Marginalization

- Given the joint distribution, we can compute distributions over smaller sets of variables through marginalization:

$$P(X{=}x) = \Sigma_{z \in dom(Z)} P(X{=}x, Z = z)$$

- This corresponds to summing out a dimension in the table.
- The new table still sums to 1. It must, since it's a probability distribution!

| *Weather* | *Temperature* | *µ(w)* |
|---|---|---|
| sunny | hot | 0.10 |
| sunny | mild | 0.20 |
| sunny | cold | 0.10 |
| cloudy | hot | 0.05 |
| cloudy | mild | 0.35 |
| cloudy | cold | 0.20 |

| *Temperature* | *µ(w)* |
|---|---|
| hot | 0.15 |
| mild | 0.55 |
| cold | 0.30 |

Alternative way to compute last entry: probabilities have to sum to 1.

# Lecture Overview

- Recap: Probability & Possible World Semantics
- Reasoning Under Uncertainty
  - Conditioning
  - Inference by Enumeration
  - Bayes Rule
  - Chain Rule

# Conditioning

- Conditioning: revise beliefs based on new observations
  - Build a probabilistic model (the joint probability distribution, JPD)
    - Takes into account all background information
    - Called the prior probability distribution
    - Denote the prior probability for hypothesis h as P(h)
  - Observe new information about the world
    - Call all information we received subsequently the evidence e
  - Integrate the two sources of information
    - to compute the conditional probability P(h|e)
    - This is also called the posterior probability of h.

- Example
  - Prior probability for having a disease (typically small)
  - Evidence: a test for the disease comes out positive
    - But diagnostic tests have false positives
  - Posterior probability: integrate prior and evidence

# Example for conditioning

- You have a prior for the joint distribution of weather and temperature, and the marginal distribution of temperature

| *Possible world* | *Weather* | *Temperature* | *μ(w)* |
|---|---|---|---|
| $w_1$ | sunny | hot | 0.10 |
| $w_2$ | sunny | mild | 0.20 |
| $w_3$ | sunny | cold | 0.10 |
| ~~$w_4$~~ | ~~cloudy~~ | ~~hot~~ | ~~0.05~~ |
| ~~$w_5$~~ | ~~cloudy~~ | ~~mild~~ | ~~0.35~~ |
| ~~$w_6$~~ | ~~cloudy~~ | ~~cold~~ | ~~0.20~~ |

| *T* | *P(T\|W=sunny)* |
|---|---|
| hot | 0.10/0.40=0.25 |
| mild | ?? |
| cold | |

0.20 0.40 0.50 0.80

- Now, you look outside and see that it's sunny
  - You are certain that you're in world $w_1$, $w_2$, or $w_3$
  - To get the conditional probability, you simply renormalize to sum to 1
  - 0.10+0.20+0.10=0.40

# Example for conditioning

- You have a prior for the joint distribution of weather and temperature, and the marginal distribution of temperature

| *Possible world* | *Weather* | *Temperature* | *μ(w)* |
|---|---|---|---|
| $w_1$ | sunny | hot | 0.10 |
| $w_2$ | sunny | mild | 0.20 |
| $w_3$ | sunny | cold | 0.10 |
| ~~$w_4$~~ | ~~cloudy~~ | ~~hot~~ | ~~0.05~~ |
| ~~$w_5$~~ | ~~cloudy~~ | ~~mild~~ | ~~0.35~~ |
| ~~$w_6$~~ | ~~cloudy~~ | ~~cold~~ | ~~0.20~~ |

| *T* | *P(T\|W=sunny)* |
|---|---|
| hot | 0.10/0.40=0.25 |
| mild | 0.20/0.40=0.50 |
| cold | 0.10/0.40=0.25 |

- Now, you look outside and see that it's sunny
  - You are certain that you're in world $w_1$, $w_2$, or $w_3$
  - To get the conditional probability, you simply renormalize to sum to 1
  - 0.10+0.20+0.10=0.40

# Semantics of Conditioning

- Evidence e ("W=sunny") rules out possible worlds incompatible with e.
  - Now we formalize what we did in the previous example

| *Possible world* | *Weather W* | *Temperature* | *$\mu(w)$* | *$\mu_e(w)$* |
|---|---|---|---|---|
| $w_1$ | sunny | hot | 0.10 | |
| $w_2$ | sunny | mild | 0.20 | |
| $w_3$ | sunny | cold | 0.10 | |
| $w_4$ | cloudy | hot | 0.05 | |
| $w_5$ | cloudy | mild | 0.35 | |
| $w_6$ | cloudy | cold | 0.20 | |

What is P(e)?

| | |
|---|---|
| 0.20 | 0.40 |
| 0.50 | 0.80 |

Recall:
e = "W=sunny"

- We represent the updated probability using a new measure, $\mu_e$, over possible worlds

$$\mu_e(w) = \begin{cases} \frac{1}{P(e)} \times \mu(w) & if \quad w \models e \\ 0 & if \quad w \not\models e \end{cases}$$

# Semantics of Conditioning

- Evidence e (“W=sunny”) rules out possible worlds incompatible with e.
  - Now we formalize what we did in the previous example

| *Possible world* | *Weather W* | *Temperature* | $\mu(w)$ | $\mu_e(w)$ |
|---|---|---|---|---|
| $w_1$ | sunny | hot | 0.10 | |
| $w_2$ | sunny | mild | 0.20 | |
| $w_3$ | sunny | cold | 0.10 | |
| $w_4$ | cloudy | hot | 0.05 | |
| $w_5$ | cloudy | mild | 0.35 | |
| $w_6$ | cloudy | cold | 0.20 | |

What is P(e)?

Marginalize out Temperature, i.e.
0.10+0.20+0.10=0.40

- We represent the updated probability using a new measure, $\mu_e$, over possible worlds

$$\mu_e(w) = \begin{cases} \frac{1}{P(e)} \times \mu(w) & if \quad w \models e \\ 0 & if \quad w \not\models e \end{cases}$$

# Semantics of Conditioning

- Evidence e ("W=sunny") rules out possible worlds incompatible with e.
  - Now we formalize what we did in the previous example

| *Possible world* | *Weather W* | *Temperature* | $\mu(w)$ | $\mu_e(w)$ |
|---|---|---|---|---|
| $w_1$ | sunny | hot | 0.10 | 0.10/0.40=0.25 |
| $w_2$ | sunny | mild | 0.20 | 0.20/0.40=0.50 |
| $w_3$ | sunny | cold | 0.10 | 0.10/0.40=0.25 |
| $w_4$ | cloudy | hot | 0.05 | 0 |
| $w_5$ | cloudy | mild | 0.35 | 0 |
| $w_6$ | cloudy | cold | 0.20 | 0 |

What is P(e)?

Marginalize out Temperature, i.e.
0.10+0.20+0.10=0.40

- We represent the updated probability using a new measure, $\mu_e$, over possible worlds

$$\mu_e(w) = \begin{cases} \frac{1}{P(e)} \times \mu(w) & if \quad w \models e \\ 0 & if \quad w \not\models e \end{cases}$$

# Conditional Probability

- P(e): Sum of probability for all worlds in which e is true
- P(h∧e): Sum of probability for all worlds in which both h and e are true
- P(h|e) = P(h∧e) / P(e)  (Only defined if P(e) > 0)

$$\mu_e(w) = \begin{cases} \frac{1}{P(e)} \times \mu(w) & \text{if} \quad w \models e \\ 0 & \text{if} \quad w \not\models e \end{cases}$$

**Definition (conditional probability)**

The conditional probability of formula h given evidence e is

$$P(h|e) = \sum_{w \models h} \mu_e(w) = \frac{1}{P(e)} \sum_{w \models h \wedge e} \mu(w) = \frac{P(h \wedge e)}{P(e)}$$

# Example for Conditional Probability

- Conditional probability distribution of Temperature given "W=sunny"
- We know $P(h|e) = \frac{P(h \wedge e)}{P(e)}$
  - E.g. $P(T = hot|W = sunny) = \frac{P(T=hot \wedge W=sunny)}{P(W=sunny)}$
  - What is P(W=sunny)?
    - Marginalize out Temperature, i.e. 0.10+0.20+0.10=0.40
- P(Temperature | W=sunny) is a new probability distribution only defined over Temperature

| *Weather W* | *Temperature T* | *P(T∧W)* |
|---|---|---|
| sunny | hot | 0.10 |
| sunny | mild | 0.20 |
| sunny | cold | 0.10 |
| cloudy | hot | 0.05 |
| cloudy | mild | 0.35 |
| cloudy | cold | 0.20 |

| *Temperature T* | *P(T\|W=sunny)* |
|---|---|
| hot | 0.10/0.40=0.25 |
| mild | 0.20/0.40=0.50 |
| cold | 0.10/0.40=0.25 |

# Lecture Overview

- Recap: Probability & Possible World Semantics
- Reasoning Under Uncertainty
  - Conditioning
  - Inference by Enumeration
  - Bayes Rule
  - Chain Rule

# Inference by Enumeration

- Great, we can compute arbitrary probabilities now!
- Given
  - Prior joint probability distribution (JPD) on set of variables X
  - specific values e for the evidence variables E (subset of X)
- We want to compute
  - posterior joint distribution of query variables Y (a subset of X) given evidence e
- Step 1: Condition to get distribution P(X|e)
- Step 2: Marginalize to get distribution P(Y|e)

# Inference by Enumeration: example

- Given P(X) as JPD below, and evidence e = “Wind=yes”
  - What is the probability it is hot? I.e., P(Temperature=hot | Wind=yes)
- Step 1: condition to get distribution P(X|e)

| *Windy W* | *Cloudy C* | *Temperature T* | *P(W, C, T)* |
|---|---|---|---|
| yes | no | hot | 0.04 |
| yes | no | mild | 0.09 |
| yes | no | cold | 0.07 |
| yes | yes | hot | 0.01 |
| yes | yes | mild | 0.10 |
| yes | yes | cold | 0.12 |
| no | no | hot | 0.06 |
| no | no | mild | 0.11 |
| no | no | cold | 0.03 |
| no | yes | hot | 0.04 |
| no | yes | mild | 0.25 |
| no | yes | cold | 0.08 |

# Inference by Enumeration: example

- Given P(X) as JPD below, and evidence e = “Wind=yes”
  - What is the probability it is hot? I.e., P(Temperature=hot | Wind=yes)
- Step 1: condition to get distribution P(X|e)

| *Windy* *W* | *Cloudy* *C* | *Temperature* *T* | *P(W, C, T)* |
|---|---|---|---|
| yes | no | hot | 0.04 |
| yes | no | mild | 0.09 |
| yes | no | cold | 0.07 |
| yes | yes | hot | 0.01 |
| yes | yes | mild | 0.10 |
| yes | yes | cold | 0.12 |
| ~~no~~ | ~~no~~ | ~~hot~~ | ~~0.06~~ |
| ~~no~~ | ~~no~~ | ~~mild~~ | ~~0.11~~ |
| ~~no~~ | ~~no~~ | ~~cold~~ | ~~0.03~~ |
| ~~no~~ | ~~yes~~ | ~~hot~~ | ~~0.04~~ |
| ~~no~~ | ~~yes~~ | ~~mild~~ | ~~0.25~~ |
| ~~no~~ | ~~yes~~ | ~~cold~~ | ~~0.08~~ |

| *Cloudy* *C* | *Temperature* *T* | *P(C, T\| W=yes)* |
|---|---|---|
| sunny | hot | |
| sunny | mild | |
| sunny | cold | |
| cloudy | hot | |
| cloudy | mild | |
| cloudy | cold | |

$$P(C = c \wedge T = t | W = yes) = \frac{P(C = c \wedge T = t \wedge W = yes)}{P(W = yes)}$$

P(W=yes) = 0.04+0.09+0.07+0.01+0.10+0.12=0.43

# Inference by Enumeration: example

- Given P(X) as JPD below, and evidence e = "Wind=yes"
  - What is the probability it is hot? I.e., P(Temperature=hot | Wind=yes)
- Step 1: condition to get distribution P(X|e)

| *Windy W* | *Cloudy C* | *Temperature T* | *P(W, C, T)* |
|---|---|---|---|
| yes | no | hot | 0.04 |
| yes | no | mild | 0.09 |
| yes | no | cold | 0.07 |
| yes | yes | hot | 0.01 |
| yes | yes | mild | 0.10 |
| yes | yes | cold | 0.12 |
| ~~no~~ | ~~no~~ | ~~hot~~ | ~~0.06~~ |
| ~~no~~ | ~~no~~ | ~~mild~~ | ~~0.11~~ |
| ~~no~~ | ~~no~~ | ~~cold~~ | ~~0.03~~ |
| ~~no~~ | ~~yes~~ | ~~hot~~ | ~~0.04~~ |
| ~~no~~ | ~~yes~~ | ~~mild~~ | ~~0.25~~ |
| ~~no~~ | ~~yes~~ | ~~cold~~ | ~~0.08~~ |

| *Cloudy C* | *Temperature T* | *P(C, T\| W=yes)* |
|---|---|---|
| sunny | hot | 0.04/0.43 ≅ 0.10 |
| sunny | mild | 0.09/0.43 ≅ 0.21 |
| sunny | cold | 0.07/0.43 ≅ 0.16 |
| cloudy | hot | 0.01/0.43 ≅ 0.02 |
| cloudy | mild | 0.10/0.43 ≅ 0.23 |
| cloudy | cold | 0.12/0.43 ≅ 0.28 |

$$P(C = c \wedge T = t | W = yes) = \frac{P(C = c \wedge T = t \wedge W = yes)}{P(W = yes)}$$

P(W=yes) = 0.04+0.09+0.07+0.01+0.10+0.12=0.43

# Inference by Enumeration: example

- Given P(X) as JPD below, and evidence e = “Wind=yes”
  - What is the probability it is hot? I.e., P(Temperature=hot | Wind=yes)
- Step 2: marginalize to get distribution P(Y|e)

| *Cloudy C* | *Temperature T* | *P(C, T\| W=yes)* |
|---|---|---|
| sunny | hot | 0.10 |
| sunny | mild | 0.21 |
| sunny | cold | 0.16 |
| cloudy | hot | 0.02 |
| cloudy | mild | 0.23 |
| cloudy | cold | 0.28 |

| *Temperature T* | *P(T\| W=yes)* |
|---|---|
| hot | 0.10+0.02 = 0.12 |
| mild | 0.21+0.23 = 0.44 |
| cold | 0.16+0.28 = 0.44 |

# Problems of Inference by Enumeration

- If we have n variables,
  and d is the size of the largest domain

- What is the space complexity to store the joint distribution?

$O(d^n)$ $O(n^d)$ $O(nd)$ $O(n+d)$

# Problems of Inference by Enumeration

- If we have n variables,
  and d is the size of the largest domain

- What is the space complexity to store the joint distribution?
  - We need to store the probability for each possible world
  - There are $O(d^n)$ possible worlds, so the space complexity is $O(d^n)$

- How do we find the numbers for $O(d^n)$ entries?
- Time complexity $O(d^n)$

- We have some of our basic tools, but
  to gain computational efficiency we need to do more
  - We will exploit (conditional) independence between variables
  - (Next week)

# Lecture Overview

- Recap: Probability & Possible World Semantics
- Reasoning Under Uncertainty
  - Conditioning
  - Inference by Enumeration
  - Bayes Rule
  - Chain Rule

# Using conditional probability

- Often you have causal knowledge:
  - For example
    - P(symptom | disease)
    - P(light is off | status of switches and switch positions)
    - P(alarm | fire)
  - In general: P(evidence e | hypothesis h)

- ... and you want to do evidential reasoning:
  - For example
    - P(disease | symptom)
    - P(status of switches | light is off and switch positions)
    - P(fire | alarm)
  - In general: P(hypothesis h | evidence e)

# Bayes rule

- By definition, we know that $P(h|e) = \frac{P(h \wedge e)}{P(e)}$

- We can rearrange terms to show:

$$P(h \wedge e) = P(h|e) \times P(e)$$

- Similarly, we can show:

$$P(e \wedge h) = P(e|h) \times P(h)$$

- Since $e \wedge h$ and $h \wedge e$ are identical, we have:

**Theorem (Bayes theorem, or Bayes rule)**

$$P(h|e) = \frac{P(e|h) \times P(h)}{P(e)}$$

# Example for Bayes rule

- On average, the alarm rings once a year
  - $P(alarm) = ?$
- If there is a fire, the alarm will almost always ring
- On average, we have a fire every 10 years
- The fire alarm rings. What is the probability there is a fire?

# Example for Bayes rule

- On average, the alarm rings once a year
  - $P(alarm) = 1/365$
- If there is a fire, the alarm will almost always ring
  - $P(alarm|fire) = 0.999$
- On average, we have a fire every 10 years
  - $P(fire) = 1/3650$
- The fire alarm rings. What is the probability there is a fire?
  - Take a few minutes to do the math!

0.999 0.9 0.0999 0.1

# Example for Bayes rule

- On average, the alarm rings once a year
  - $P(alarm) = 1/365$
- If there is a fire, the alarm will almost always ring
  - $P(alarm|fire) = 0.999$
- On average, we have a fire every 10 years
  - $P(fire) = 1/3650$
- The fire alarm rings. What is the probability there is a fire?
- $P(fire|\text{alarm}) = \frac{P(alarm|fire) \times P(fire)}{P(alarm)} = \frac{0.999 \times 1/3650}{1/365} = 0.0999$
  - Even though the alarm rings the chance for a fire is only about 10%!

# Lecture Overview

- Recap: Probability & Possible World Semantics
- Reasoning Under Uncertainty
  - Conditioning
  - Bayes Rule
  - Inference by Enumeration
  - Chain Rule

# Product Rule

- By definition, we know that

$$P(f_2|f_1) = \frac{P(f_2 \wedge f_1)}{P(f_1)}$$

- We can rewrite this to

$$P(f_2 \wedge f_1) = P(f_2|f_1) \times P(f_1)$$

- In general:

**Theorem (Product Rule)**

$$P(f_n \wedge \cdots \wedge fi_{+1} \wedge f_{\mathrm{i}} \wedge \cdots \wedge f_1) = P(f_n \wedge \cdots \wedge fi_{+1}|f_{\mathrm{i}} \wedge \cdots \wedge f_1) \times P(f_{\mathrm{i}} \wedge \cdots \wedge f_1)$$

# Chain Rule

- We know

$$P(f_2 \wedge f_1) = P(f_2 | f_1) \times P(f_1)$$

- In general:

$$P(f_n \wedge f_{n-1} \wedge \cdots \wedge f_1)$$

$$= P(f_n | f_{n-1} \wedge \cdots \wedge f_1) \times P(f_{n-1} \wedge \cdots \wedge f_1)$$

$$= P(f_n | f_{n-1} \wedge \cdots \wedge f_1) \times P(f_{n-1} | f_{n-2} \wedge \cdots \wedge f_1) \times P(f_{n-2} \wedge \cdots \wedge f_1)$$

= ...

$$= \prod_{i=1}^{n} P(f_i | f_{i-1} \wedge \cdots \wedge f_1)$$

**Theorem (Chain Rule)**

$$P(f_n \wedge \cdots \wedge f_1) = \prod_{i=1}^{n} P(fi | f_{i-1} \wedge \cdots \wedge f_1)$$

# Why does the chain rule help us?

- We can simplify some terms
  - For example, how about P(Weather | PriceOfOil) ?
    - Weather in Vancouver is independent of the price of oil:

$$P(Weather|PriceOfOil) = P(Weather)$$

- Under independence, we gain compactness
  - We can represent the JPD as a product of marginal distributions
  - For example: P(Weather,PriceOfOil) = P(Weather) × P(PriceOfOil)
  - But not all variables are independent

$$P(Weather|Temperature) \neq P(Weather)$$

    - More about (conditional) independence next week

# Learning Goals For Today's Class

- Prove the formula to compute conditional probability P(h|e)
- Use inference by enumeration
  - to compute joint posterior probability distributions over any subset of variables given evidence
- Derive and use Bayes Rule
- Derive the Chain Rule

---

- Marginalization, conditioning and Bayes rule are crucial
  - They are core to reasoning under uncertainty
  - Be sure you understand them and be able to use them!
- First question of assignment 4 available on WebCT
  - Simple application of Bayes rule
  - Do it as an exercise before next class