Finish Logics

Start Stochastic Environments

Computer Science cpsc322, Lecture 8

Textbook Chpt 5.2 & Chpt 12

Chpt 6.1

May, 31, 2012

Lecture Overview

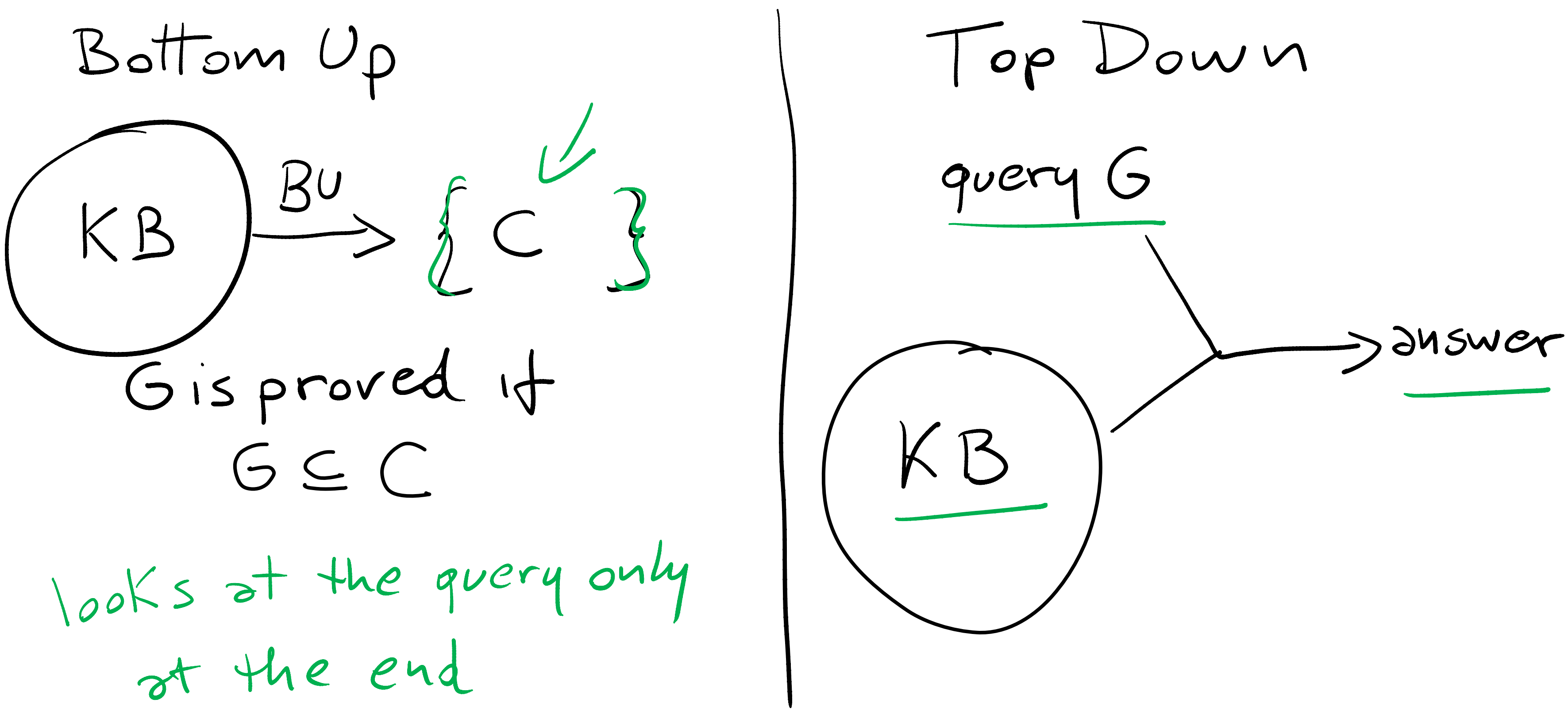
- Finish Logics
  - Recap Top Down + TD as Search
  - Datalog

Start Stochastic Environments

- Intro to Probability
  - Semantics of Probability
  - Marginalization
  - Conditional Probability and Chain Rule
  - Bayes' Rule and Independence

Top-down Ground Proof Procedure

Key Idea: search backward from a query *G* to determine if it can be derived from *KB*.



Top-down Proof Procedure: Basic elements

Notation: An answer clause is of the form:

$yes \leftarrow a_1 \wedge a_2 \wedge \ldots \wedge a_m$

Express query as an answer clause

(e.g., query $a_1 \wedge a_2 \wedge \ldots \wedge a_m$)

$yes \leftarrow$ $a_1 \wedge \ldots \wedge a_m$

Rule of inference (called SLD Resolution)

Given an answer clause of the form:

$yes \leftarrow a_1 \wedge a_2 \wedge \ldots \wedge a_m$

and the clause: in KB

$a_i \leftarrow b_1 \wedge b_2 \wedge \ldots \wedge b_p$

$a_1 \leftarrow$

a_i

You can generate the answer clause

$yes \leftarrow a_1 \wedge \ldots \wedge a_{i-1} \wedge b_1 \wedge b_2 \wedge \ldots \wedge b_p \wedge a_{i+1} \wedge \ldots \wedge a_m$

- **Successful Derivation**: When by applying the inference rule you obtain the answer clause *yes* ← .

$a \leftarrow e \wedge f.$	$a \leftarrow b \wedge c.$	$b \leftarrow k \wedge f.$
$c \leftarrow e.$	$d \leftarrow k.$	$e.$
$f \leftarrow j \wedge e.$	$f \leftarrow c.$	$j \leftarrow c.$

Query: a `(two ways)`

$yes \leftarrow a.$

$yes \leftarrow a.$

Lecture Overview

- Finish Logics
  - Recap Top Down + TD as Search
  - Datalog

Start Stochastic Environments

- Intro to Probability
  - Semantics of Probability
  - Marginalization
  - Conditional Probability and Chain Rule
  - Bayes' Rule and Independence

Systematic Search in different R&R systems

Constraint Satisfaction (Problems): ✓

- State: assignments of values to a subset of the variables
- Successor function: assign values to a "free" variable
- Goal test: set of constraints
- Solution: possible world that satisfies the constraints
- Heuristic function: *none (all solutions at the same distance from start)*

Planning (forward) : ✓

- State possible world
- Successor function states resulting from valid actions
- Goal test assignment to subset of vars
- Solution sequence of actions
- Heuristic function empty-delete-list (solve simplified problem)

Logical Inference (top Down)

- State answer clause   yes ←
- Successor function states resulting from substituting one atom with all the clauses of which it is the head
- Goal test empty answer clause yes ←
- Solution start state
- Heuristic function ………………… ✓ see next slide

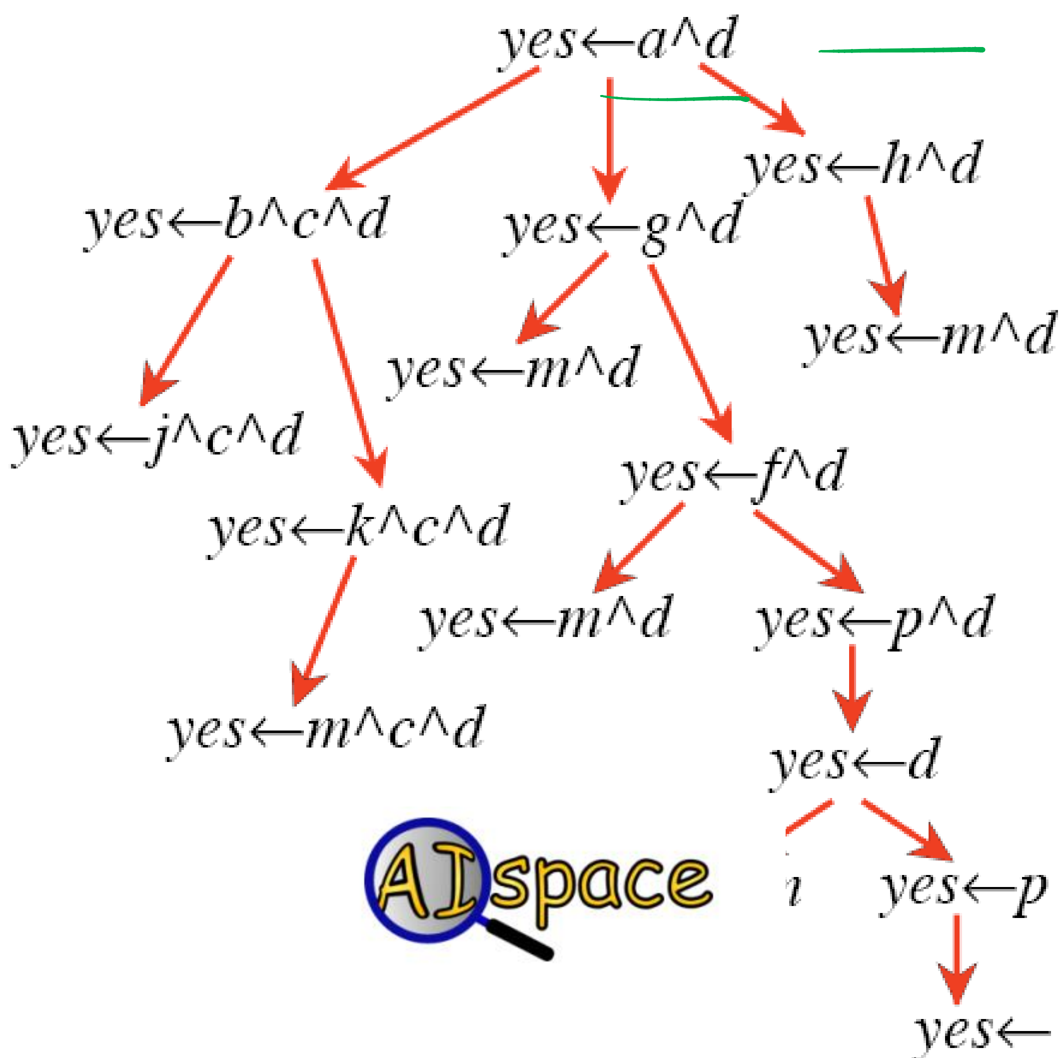
start state: query as an answer clause

Search Graph

KB

$a \leftarrow b \wedge c.$	$a \leftarrow g.$
$a \leftarrow h.$	$b \leftarrow j.$
$b \leftarrow k.$	$d \leftarrow m.$
$d \leftarrow p.$	$f \leftarrow m.$
$f \leftarrow p.$	$g \leftarrow m.$
$g \leftarrow f.$	$k \leftarrow m.$
$h \leftarrow m.$	$p.$

Prove: $? \leftarrow a \wedge d.$



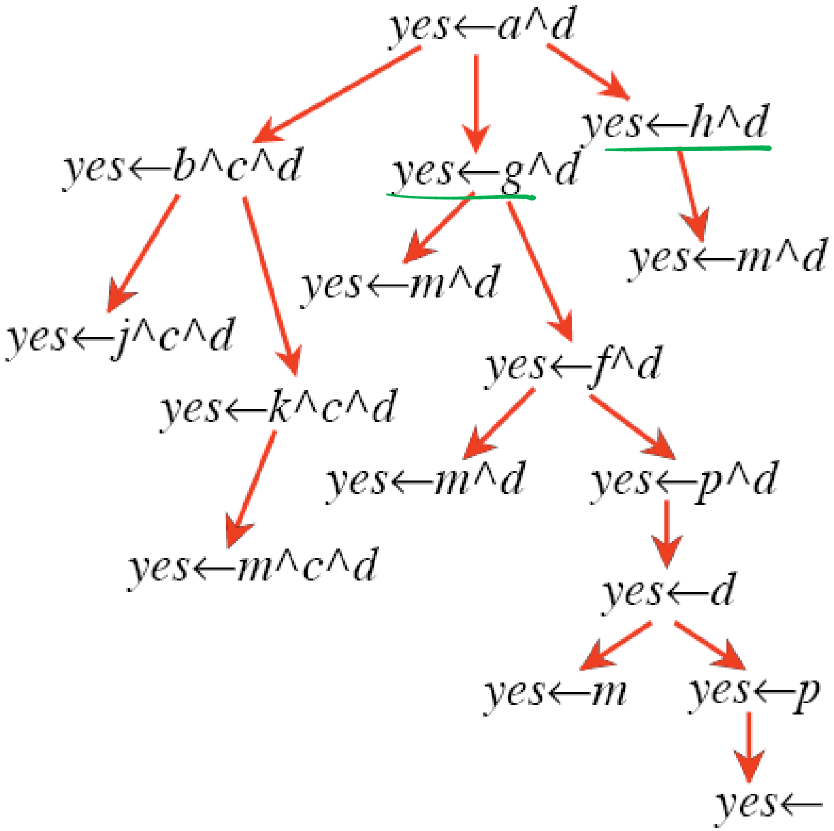
Heuristics?

Search Graph

KB

$a \leftarrow b \wedge c.$	$a \leftarrow g.$
$a \leftarrow h.$	$b \leftarrow j.$
$b \leftarrow k.$	$d \leftarrow m.$
$d \leftarrow p.$	$f \leftarrow m.$
$f \leftarrow p.$	$g \leftarrow m.$
$g \leftarrow f.$	$k \leftarrow m.$
$h \leftarrow m.$	$p.$

Prove: $? \leftarrow a \wedge d.$



Possible Heuristic?

Number of atoms in the answer clause

Admissible?

Yes No

Standard Search vs. Specific R&R systems

Constraint Satisfaction (Problems):

- State: assignments of values to a subset of the variables
- Successor function: assign values to a “free” variable
- Goal test: set of constraints
- Solution: possible world that satisfies the constraints
- Heuristic function: *none (all solutions at the same distance from start)*

Planning :

- State possible world
- Successor function states resulting from valid actions
- Goal test assignment to subset of vars
- Solution sequence of actions
- Heuristic function empty-delete-list (solve simplified problem)

Logical Inference

- State answer clause
- Successor function states resulting from substituting one atom with all the clauses of which it is the head
- Goal test empty answer clause
- Solution start state
- Heuristic function number of atoms in given state

Lecture Overview

- Finish Logics
  - Recap Top Down + TD as Search
  - Datalog

Start Stochastic Environments

- Intro to Probability
  - Semantics of Probability
  - Marginalization
  - Conditional Probability and Chain Rule
  - Bayes' Rule and Independence

Representation and Reasoning in Complex domains

- In complex domains expressing knowledge with **propositions** can be quite limiting

- It is often natural to consider **individuals** and their **properties**

p *up_s*$_2$
q *up_s*$_3$
r *ok_cb*$_1$
s *ok_cb*$_2$
live_w$_1$
connected_w$_1$*_w*$_2$

up(s_2 *)*
up(s_3 *)*
ok(cb_1 *)*
ok(cb_2 *)*
live(w_1*)*
connected(w_1 , w_2 *)*

There is no notion that the system can reason about

up_s$_2$
up_s$_3$

up are about the same property

live_w$_1$
connected_w$_1$*_w*$_2$

w_1 are about the same individual

What do we gain….

By breaking propositions into relations applied to individuals?

- Express **knowledge** that **holds for set of individuals** (by introducing variables)

  live(W) <- connected_to(W,W1) ∧ live(W1) ∧ wire(W) ∧ wire(W1).

- We can **ask generic queries** (i.e., containing vars)
  variabless

  ? connected_to(W, w_1*)*

Datalog vs PDCL (better with colors)

First Order Logic

$\forall X \exists Y p(X,Y) \Leftrightarrow \neg q(Y)$

$p(a_1, a_2)$

$\neg q(a_5)$

Datalog

$p(X) \leftarrow q(X) \wedge r(X,Y)$

$r(X,Y) \leftarrow s(Y)$

$s(a_1), q(a_2)$

Propositional Logic

$\neg(p \vee q) \rightarrow (r \wedge s \wedge f),$

p, r

PDCL

$p \leftarrow s \wedge f$

$r \leftarrow s \wedge q \wedge p$

r

p

Datalog: a relational rule language

Datalog expands the syntax of PDCL….

A variable is a symbol starting with an upper case letter

Examples: X, Y

A constant is a symbol starting with lower-case letter or a sequence of digits.

Examples: alan, w1

A term is either a variable or a constant.

Examples: X, Y, alan, w1

A predicate symbol is a symbol starting with a lower-case letter.

Examples: live, connected, part-of, in

Datalog Syntax (cont'd)

An atom is a symbol of the form p or $p(t_1 \dots t_n)$ where p is a predicate symbol and t_i are terms

Examples: sunny, in(alan,X)

A definite clause is either an atom (a fact) or of the form:

$$h \leftarrow b_1 \wedge \dots \wedge b_m$$

where h and the b_i are atoms (Read this as ``h if b.")

Example: in(X,Z) ← in(X,Y) ∧ part-of(Y,Z)

A knowledge base is a set of definite clauses

Datalog: Top Down Proof Procedure

in(alan, r123).
part_of(r123,cs_building).
in(X,Y) ← part_of(Z,Y) & in(X,Z).

- Extension of Top-Down procedure for PDCL.
  How do we deal with variables?
  - Idea:
    - Find a clause with head that matches the query
    - Substitute variables in the clause with their matching constants
  - Example:

Query: yes ← in(alan, cs_building).

↓

in(X,Y) ← part_of(Z,Y) & in(X,Z).
with Y = cs_building
X = alan

yes ← part_of(Z,cs_building), in(alan, Z).

- We will not cover the formal details of this process, called ***unification***. See P&M Section 12.4.2, p. 511 for the details.

Example proof of a Datalog query

in(alan, r123).

part_of(r123,cs_building).

in(X,Y) ← part_of(Z,Y) & in(X,Z).

Query: yes ← in(alan, cs_building).

Using clause: in(X,Y) ← part_of(Z,Y) & in(X,Z), with Y = cs_building X = alan

yes ← part_of(Z,cs_building), in(alan, Z).

Using clause: part_of(r123,cs_building) with Z = r123

yes ← in(alan, r123).

Using clause: in(alan, r123).

Using clause: in(X,Y) ← part_of(Z,Y) & in(X,Z). With X = alan Y = r123

yes ←.

yes ← part_of(Z, r123), in(alan, Z).

No clause with matching head: part_of(Z,r123).

fail

Tracing Datalog proofs in AIspace

- You can trace the example from the last slide in the AIspace Deduction Applet at http://aispace.org/deduction/ using file *ex-Datalog* available in course schedule

- Question 4 of assignment 3 asks you to use this applet

Datalog: queries with variables

in(alan, r123).
part_of(r123,cs_building).
in(X,Y) ← part_of(Z,Y) & in(X,Z).

Query: in(alan, X1).

yes(X1) ← in(alan, X1).

What would the answer(s) be?

Datalog: queries with variables

in(alan, r123).
part_of(r123,cs_building).
in(X,Y) ← part_of(Z,Y) & in(X,Z).

Query: in(alan, X1).

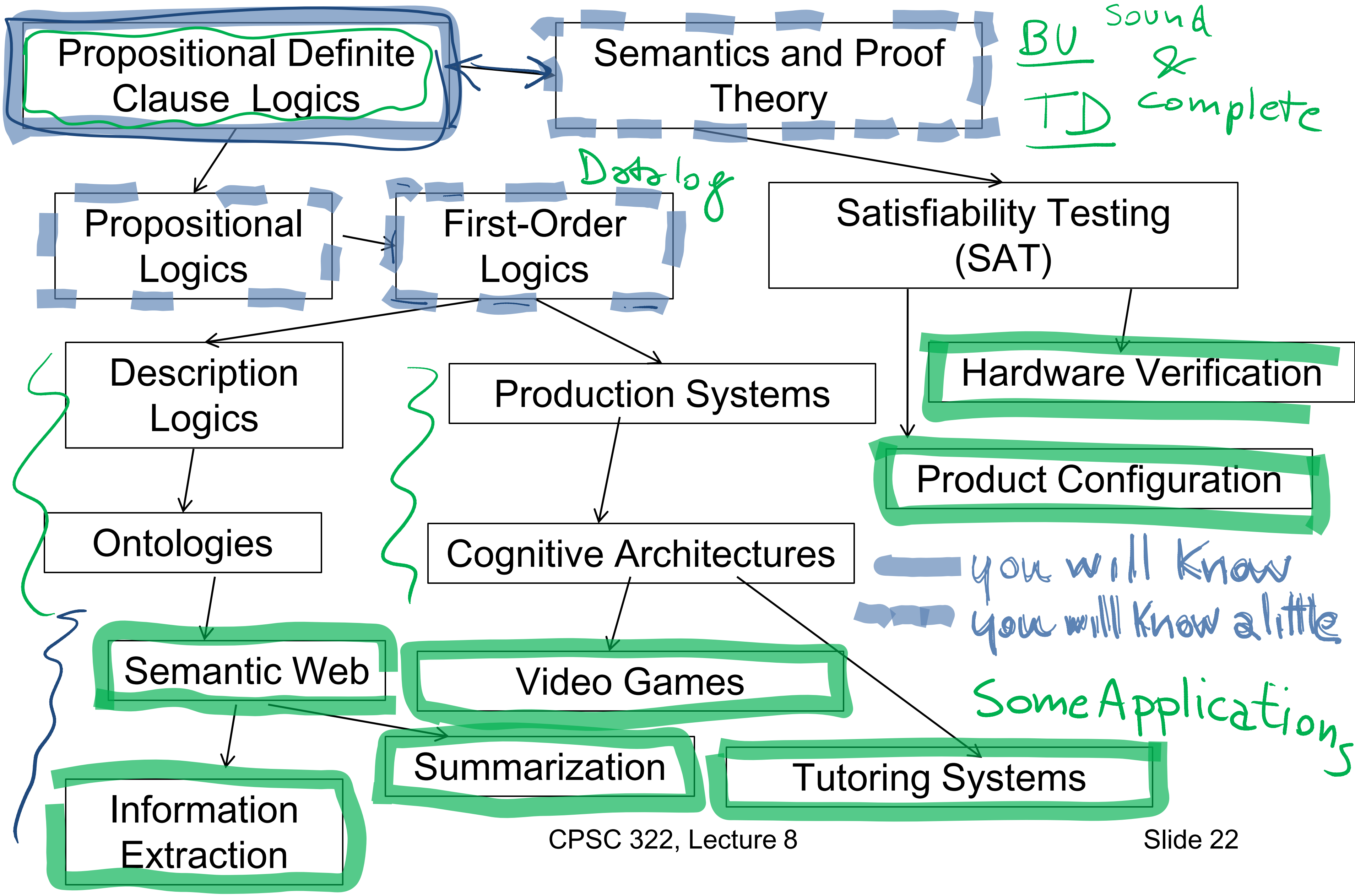
yes(X1) ← in(alan, X1).

What would the answer(s) be?

yes(r123).
yes(cs_building).

Again, you can trace the SLD derivation for this query
in the AIspace Deduction Applet

Logics in AI: Similar slide to the one for planning



Learning Goals for today's class

You can:

- Define/read/write/trace/debug the **TopDown** proof procedure (as a **search** problem)
- Represent simple domains in **Datalog**
- Apply **TopDown** proof procedure in **Datalog**

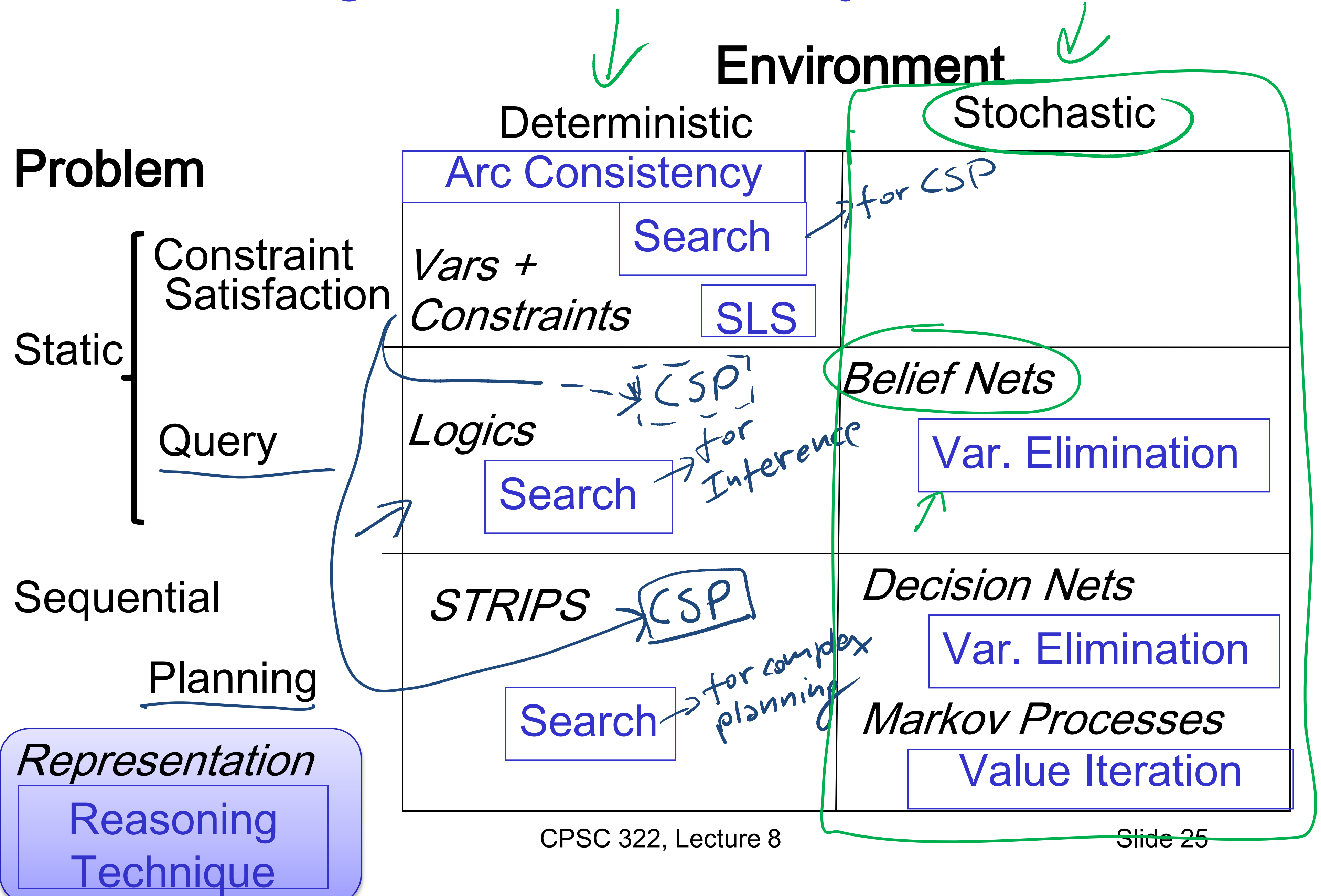
Lecture Overview

- Finish Logics
  - Recap Top Down + TD as Search
  - Datalog

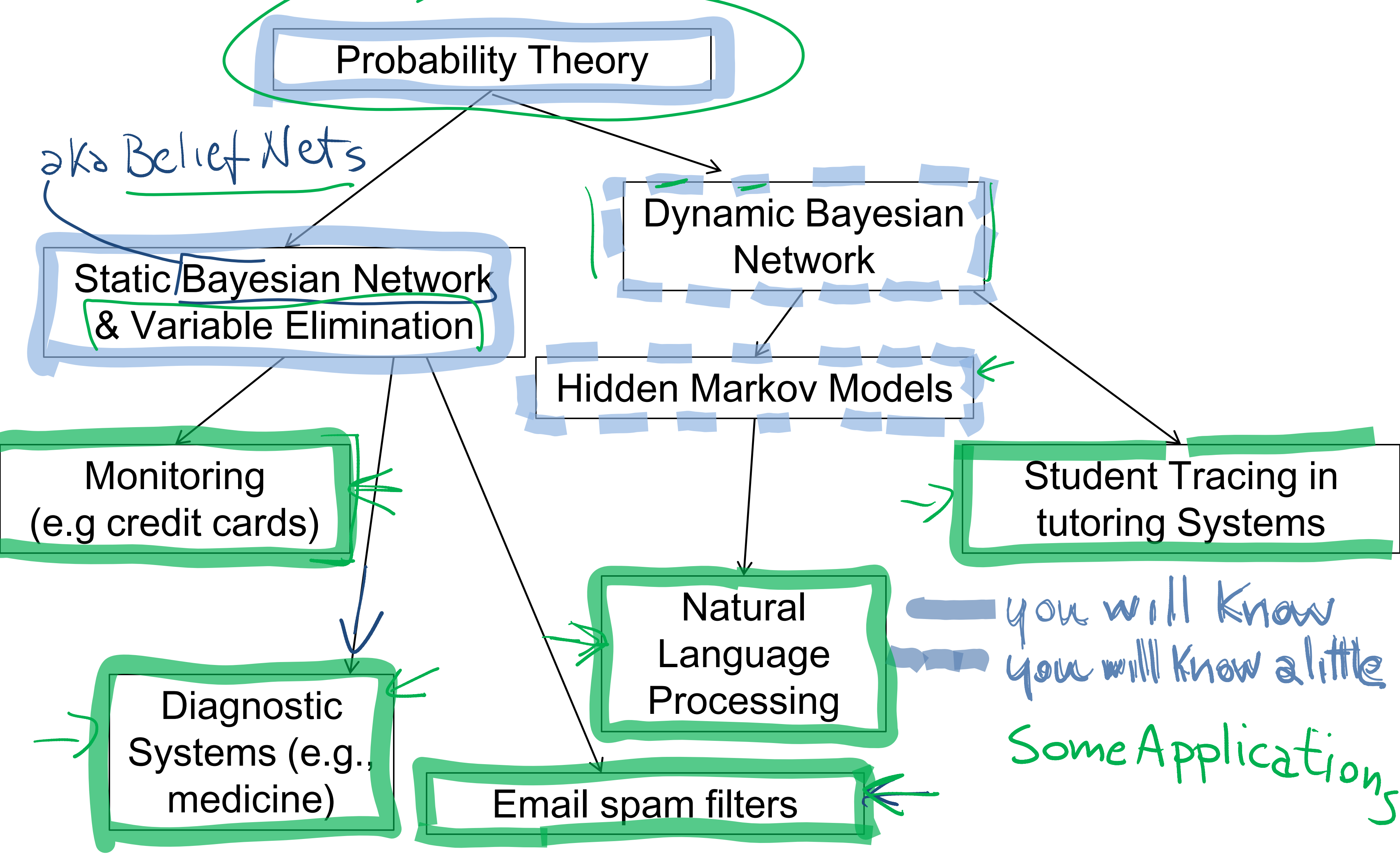
Start Stochastic Environments

- Intro to Probability
  - Semantics of Probability
  - Marginalization
  - Conditional Probability and Chain Rule
  - Bayes' Rule and Independence

Big Picture: R&R systems



Answering Query under Uncertainty



Intro to Probability (Motivation)

- *Will it rain in 10 days? Was it raining 98 days ago?*
- *Right now, how many people are in this room? in this building (DMP)? At UBC? ….Yesterday?*

- AI agents (and humans ☹) are not omniscient (know everything)

  they are ignorant

- And the problem is not only predicting the future or "remembering" the past

  also current state

Intro to Probability (Key points)

- Are agents all ignorant/uncertain to the same degree? No

  it is subjective

- Should an agent act only when it is certain about relevant knowledge?
- (not acting usually has implications)

- So agents need to **represent and reason about their ignorance/ uncertainty**

Probability as a formal measure of uncertainty/ignorance

- Belief in a proposition *f* (e.g., *it is raining outside, there are 31 people in this room*) can be measured in terms of a number between 0 and 1 – this is the probability of *f*
  - The probability *f* is 0 means that *f* is believed to be
    definitely false
  - The probability *f* is 1 means that *f* is believed to be
    definitely true
  - Using 0 and 1 is purely a convention.

Random Variables

- A random variable is a **variable** like the ones we have seen in **CSP** and **Planning**, but the agent can be **uncertain about its value**.
- As usual
  - The domain of a random variable *X*, written *dom(X)*, is the set of values *X* can take
  - values are mutually exclusive and exhaustive

Examples (Boolean and discrete)

outside Raining

T F

#-of-people-rm

$[0, 10^3]$

Random Variables (cont')

- A tuple of random variables <$X_1, \ldots, X_n$> is a complex random variable with domain..

  dom(X_1) x . . . x dom(X_n)

- Assignment $X=x$ means X has value x

  outsideRaining = T

- A proposition is a Boolean formula made from assignments of values to variables

Examples

outsideRaining = T ∧ (AND) / ∨ (OR) #people-rm = 47

Lecture Overview

- Finish Logics
  - Recap Top Down + TD as Search
  - Datalog

Start Stochastic Environments

- Intro to Probability
  - **Semantics of Probability**
  - Marginalization
  - Conditional Probability and Chain Rule
  - Bayes' Rule and Independence

Possible Worlds

- A possible world specifies an assignment to each random variable

E.g., if we model only two Boolean variables *Cavity* and *Toothache*, then there are 4 distinct possible worlds:

w_1 *Cavity = T* $\wedge$ *Toothache = T*
w_2 *Cavity = T* $\wedge$ *Toothache = F*
w_3 *Cavity = F* $\wedge$ *Toothache = T*
w_4 *Cavity = T* $\wedge$ *Toothache = T*

cavity	*toothache*
T	T
T	F
F	T
F	F

As usual, possible worlds are mutually exclusive and exhaustive

$w \models X{=}x$ means variable X is assigned value x in world w

$w_3 \models$ Cavity F $\qquad$ $w_4 \not\models$ Toothache = F

Semantics of Probability

- The belief of being in each possible world *w* can be expressed as a probability *μ(w)*
- For sure, I must be in one of them……so

set of all possible worlds $\sum_{w \in W} \mu(w) = 1$

μ(w) for possible worlds generated by three Boolean variables: ***cavity***, ***toothache***, ***catch*** *(the probe caches in the tooth)*

cavity	*toothache*	*catch*	*μ(w)*
T	T	T	.108
T	T	F	.012
T	F	T	.072
T	F	F	.008
F	T	T	.016
F	T	F	.064
F	F	T	.144
F	F	F	.576

$\sum = 1$

Probability of proposition

- What is the **probability of a proposition *f***?

cavity	*toothache*	*catch*	$\mu(w)$
T	T	T	.108
T	T	F	.012
T	F	T	.072
T	F	F	.008
F	T	T	.016
F	T	F	.064
F	F	T	.144
F	F	F	.576

equivalent, only different notation

	toothache		¬ *toothache*	
	catch	¬ *catch*	*catch*	¬ *catch*
cavity	.108	.012	.072	.008
¬ *cavity*	.016	.064	.144	.576

P(toothache=F) = .8

For any *f*, sum the prob. of the worlds where it is true:

$$P(f)=\Sigma_{w \models f}\, \mu(w)$$

Ex: P(*toothache = T*) = .2

Probability of proposition

- What is the **probability of a proposition *f*** ?

cavity	*toothache*	*catch*	$\mu(w)$
T	T	T	.108
T	T	F	.012
T	F	T	.072
T	F	F	.008
F	T	T	.016
F	T	F	.064
F	F	T	.144
F	F	F	.576

	toothache		¬ *toothache*	
	catch	¬ *catch*	*catch*	¬ *catch*
cavity	.108	.012	.072	.008
¬ *cavity*	.016	.064	.144	.576

For any *f*, sum the prob. of the worlds where it is true:

$$P(f)=\sum_{w \models f} \mu(w)$$

P(cavity=T and toothache=F) = .08

Probability of proposition

- What is the **probability of a proposition** *f*?

cavity	*toothache*	*catch*	$\mu(w)$
T	T	T	.108
T	T	F	.012
T	F	T	.072
T	F	F	.008
F	T	T	.016
F	T	F	.064
F	F	T	.144
F	F	F	.576

	toothache		¬ *toothache*	
	catch	¬ *catch*	*catch*	¬ *catch*
cavity	.108	.012	.072	.008
¬ *cavity*	.016	.064	.144	.576

For any *f*, sum the prob. of the worlds where it is true:

$$P(f)=\Sigma_{w \models f}\, \mu(w)$$

P(cavity or *toothache*) = 0.108 + 0.012 + 0.016 + 0.064 + + 0.072+0.08 = 0.28

= 1 − (.144 + .576)

One more example

- Weather, with domain {sunny, cloudy)
- Temperature, with domain {hot, mild, cold}
- There are now 6 possible worlds:
- What's the probability of it being cloudy or cold?

1 0.6 0.3 0.7

Weather	*Temperature*	$\mu(w)$
sunny	hot	0.10
sunny	mild	0.20
sunny	cold	0.10
cloudy	hot	0.05
cloudy	mild	0.35
cloudy	cold	0.20

- Remember
  - The probability of proposition f is defined by: $P(f) = \Sigma_{w \models f} \mu(w)$
  - sum of the probabilities of the worlds w in which f is true

One more example

- Weather, with domain {sunny, cloudy)
- Temperature, with domain {hot, mild, cold}
  - There are now 6 possible worlds:
  - What's the probability of it being cloudy or cold?
  - μ(w3) + μ(w4) + μ(w5) + μ(w6) = 0.7

	Weather	*Temperature*	*µ(w)*
w1	sunny	hot	0.10
w2	sunny	mild	0.20
w3	sunny	cold	0.10
w4	cloudy	hot	0.05
w5	cloudy	mild	0.35
w6	cloudy	cold	0.20

- Remember
  - The probability of proposition f is defined by: $P(f) = \Sigma_{w \models f} \mu(w)$
  - sum of the probabilities of the worlds w in which f is true

Probability Distributions

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

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

dom(cavity) = [T, F]

cavity? X

T → .2 P(cavity=T)

F → .8 P(cavity=F)

	toothache		*¬ toothache*		
	catch	*¬ catch*	*catch*	*¬ catch*	
cavity	**.108**	**.012**	**.072**	**.008**	.2
¬ cavity	**.016**	**.064**	**.144**	**.576**	.8

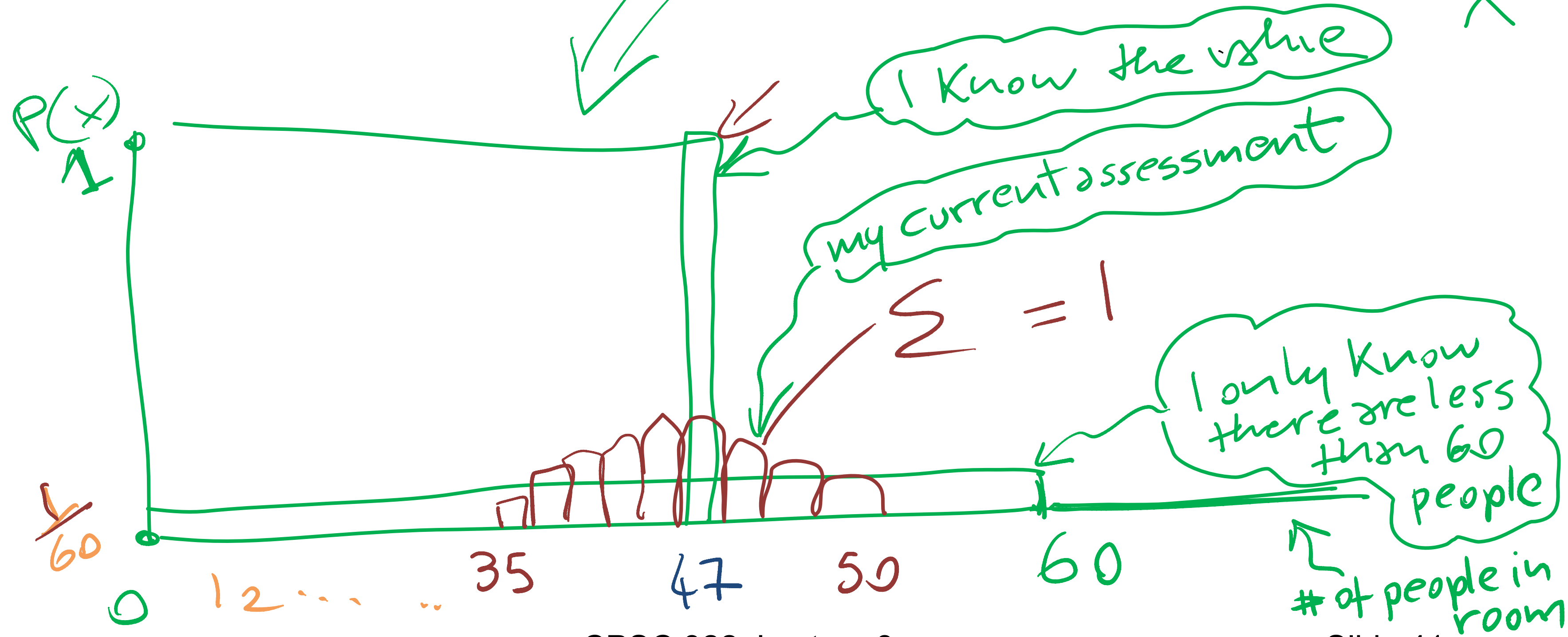
cavity	*toothache*	*catch*	*µ(w)*
T	T	T	.108
T	T	F	.012
T	F	T	.072
T	F	F	.008
F	T	T	.016
F	T	F	.064
F	F	T	.144
F	F	F	.576

Probability distribution (non binary)

- A probability distribution P on a random variable X is a function *dom(X)* - > *[0,1]* such that

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

- Number of people in this room at this time



Joint Probability Distributions

- When we have multiple random variables, their joint distribution is a probability distribution over the variable Cartesian product
  - E.g., P(<$X_1, \ldots, X_n$>)
  - Think of a joint distribution over n variables as an n-dimensional table
  - Each entry, indexed by $X_1 = x_1, \ldots, X_n = x_n$ corresponds to $P(X_1 = x_1 \wedge \ldots \wedge X_n = x_n)$
  - The sum of entries across the whole table is 1

	toothache		¬ *toothache*	
	catch	¬ *catch*	*catch*	¬ *catch*
cavity	.108	.012	.072	.008
¬ *cavity*	.016	.064	.144	.576

Question

- If you have the joint of n variables. Can you compute the probability distribution for each variable?

yes you can compute the prob. of any proposition in $X_1 \ldots X_n$

Learning Goals for today's class

You can:

- Define and give examples of **random variables**, their domains and probability distributions.
- Calculate the **probability of a proposition f** given *μ(w)* for the set of possible worlds.
- Define a **joint probability distribution**

Recap: Possible World Semantics for Probabilities

Probability is a formal measure of subjective uncertainty.

Random variable and probability distribution

X $\quad dom(X) = \{x_1, x_2, x_3\}$

$x_1 \rightarrow P(x_1)$, $x_2 \rightarrow P(x_2)$, $x_3 \rightarrow P(x_3)$ $\quad \Sigma = 1$

- Model Environment with a set of random vars

$X \; Y \; Z$ binary $\quad 8$

$$\sum_{w \in W} \mu(w) = 1$$

- Probability of a proposition *f*

formula $\quad X = T \wedge Z = F$

$$P(f) = \sum_{w \models f} \mu(w)$$

Lecture Overview

- Finish Logics
  - Recap Top Down + TD as Search
  - Datalog

Start Stochastic Environments

- Intro to Probability
  - Semantics of Probability
  - **Marginalization**
  - Conditional Probability and Chain Rule
  - Bayes' Rule and Independence

Joint Distribution and Marginalization

cavity	*toothache*	*catch*	*µ(w)*
T	T	T	.108
T	T	F	.012
T	F	T	.072
T	F	F	.008
F	T	T	.016
F	T	F	.064
F	F	T	.144
F	F	F	.576

$P(cavity, toothache, catch)$

Given a joint distribution, e.g. *P(X,Y, Z)* we can compute distributions over any smaller sets of variables

$$P(X,Y) = \sum_{z \in dom(Z)} P(X,Y,Z=z)$$

	toothache		*¬ toothache*	
	catch	*¬ catch*	*catch*	*¬ catch*
cavity	.108	.012	.072	.008
¬ cavity	.016	.064	.144	.576

cavity	*toothache*	*P(cavity , toothache)*
T	T	.12
T	F	.08
F	T	.08
F	F	.72

Why is it called Marginalization?

P(X, Y)

cavity	toothache	P(cavity , toothache)
T	T	.12
T	F	.08
F	T	.08
F	F	.72

$$P(X) = \sum_{y \in dom(Y)} P(X, Y = y)$$

P(cavity)

	Toothache = T	Toothache = F
Cavity = T	.12	.08
Cavity = F	.08	.72

.2

.8

.2 .8

P(toothache)

Marginalization

- Can we also marginalize over more than one variable at once?
- E.g. go from P(Wind, Weather, Temperature) to P (Weather)?

Wind	*Weather*	*Temperature*	*μ(w)*
yes	sunny	hot	0.04
yes	sunny	mild	0.09
yes	sunny	cold	0.07
yes	cloudy	hot	0.01
yes	cloudy	mild	0.10
yes	cloudy	cold	0.12
no	sunny	hot	0.06
no	sunny	mild	0.11
no	sunny	cold	0.03
no	cloudy	hot	0.04
no	cloudy	mild	0.25
no	cloudy	cold	0.08

i.e., Marginalization over Temperature and Wind

Weather	*μ(w)*
sunny	
cloudy	

Marginalization

- Can we also marginalize over more than one variable at once?
- E.g. go from P(Wind, Weather, temperature) to P (Weather)?

Wind	*Weather*	*Temperature*	*µ(w)*
yes	sunny	hot	0.04
yes	sunny	mild	0.09
yes	sunny	cold	0.07
yes	cloudy	hot	0.01
yes	cloudy	mild	0.10
yes	cloudy	cold	0.12
no	sunny	hot	0.06
no	sunny	mild	0.11
no	sunny	cold	0.03
no	cloudy	hot	0.04
no	cloudy	mild	0.25
no	cloudy	cold	0.08

i.e., Marginalization over Temperature and Wind

Weather	*µ(w)*
sunny	???
cloudy	

How can we compute P(Weather = sunny)?

Marginalization

- We can also marginalize over more than one variable at once

$$P(X=x) = \Sigma_{z_1 \in dom(Z_1),\ldots,\, z_n \in dom(Z_n)} P(X=x, Z_1 = z_1, \ldots, Z_n = z_n)$$

Wind	*Weather*	*Temperature*	*µ(w)*
yes	sunny	hot	0.04
yes	sunny	mild	0.09
yes	sunny	cold	0.07
yes	cloudy	hot	0.01
yes	cloudy	mild	0.10
yes	cloudy	cold	0.12
no	sunny	hot	0.06
no	sunny	mild	0.11
no	sunny	cold	0.03
no	cloudy	hot	0.04
no	cloudy	mild	0.25
no	cloudy	cold	0.08

i.e., Marginalization over Temperature and Wind

Weather	*µ(w)*
sunny	0.40
cloudy	

Lecture Overview

- Finish Logics
  - Recap Top Down + TD as Search
  - Datalog

Start Stochastic Environments

- Intro to Probability
  - Semantics of Probability
  - Marginalization
  - **Conditional Probability and Chain Rule**
  - Bayes' Rule and Independence

Conditioning (Conditional Probability)

- We **model our environment** with a **set of random variables**.
- Assume have **the joint**, we can compute the probability .... any formula
- Are we done with reasoning under uncertainty?
- What can happen?
- Think of a patient showing up at the dentist office. *Does she have a cavity?*

Conditioning (Conditional Probability)

- Probabilistic conditioning specifies **how to revise beliefs based on new information.**
- You build a probabilistic model (for now the joint) taking all background information into account. This gives the prior probability.
- All other information must be conditioned on.
- If evidence *e* is all of the information obtained subsequently, the conditional probability $P(h|e)$ of h given e is the posterior probability of h.

Conditioning Example

- Prior probability of having a cavity

P(cavity = T)

- Should be revised if you know that there is toothache

P(cavity = T | toothache = T)

- It should be revised again if you were informed that the probe did not catch anything

P(cavity =T | toothache = T, catch = F)

- What about ?

P(cavity = T | sunny = T)

How can we compute P(h|e)

- What happens in term of possible worlds if we know the value of a random var (or a set of random vars)?
- Some worlds are ruled out. The other become …. more likely

$\Sigma = P(e) = .2$

cavity	toothache	catch	$\mu(w)$	$\mu_e(w)$
T	T	T	.108	.54
T	T	F	.012	.06
T	F	T	.072	.36
T	F	F	.008	.04
F	T	T	.016	0
F	T	F	.064	0
F	F	T	.144	0
F	F	F	.576	0

$$e = (cavity = T)$$

$$\mu_e(w) = \frac{\mu(w)}{P(e)}$$

$w \models e$

Semantics of Conditional Probability

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

- The conditional probability of formula ***h*** given evidence ***e*** is A

$$P(h \mid e) = \sum_{w \models h} \mu_e(w) = \sum \frac{1}{P(e)} \times \mu(w) = \frac{1}{P(e)} \sum \mu(w)$$

$w \models h \wedge e$ $\quad$ $w \models h \wedge e$

$$= \frac{P(h \wedge e)}{P(e)} \quad B$$

Semantics of Conditional Prob.: Example

$e = (cavity = T)$

cavity	*toothache*	*catch*	$\mu(w)$	$\mu_e(w)$
T	T	T	.108	.54
T	T	F	.012	.06
T	F	T	.072	.36
T	F	F	.008	.04
F	T	T	.016	0
F	T	F	.064	0
F	F	T	.144	0
F	F	F	.576	0

$P(h \mid e) = P(toothache = T \mid cavity = T) =$

(A) $\sum_{w \models h} \mu_e(w) = .6$

(B) $\frac{P(h \wedge e)}{P(e)}$

$\frac{.12}{.2} = .6$

Conditional Probability among Random Variables

$$P(X \mid Y) = P(X , Y) / P(Y)$$

$$P(X \mid Y) = P(toothache \mid cavity)$$
$$= P(toothache \wedge cavity) / P(cavity)$$

	Toothache = T	Toothache = F
Cavity = T	.12	.08
Cavity = F	.08	.72

	Toothache = T	Toothache = F
Cavity = T	.6	.4
Cavity = F	.1	.9

Product Rule

Definition of conditional probability:

- $P(X_1 \mid X_2) = P(X_1 , X_2) / P(X_2)$

Product rule gives an alternative, more intuitive formulation:

- $P(X_1 , X_2) = P(X_2) P(X_1 \mid X_2) = P(X_1) P(X_2 \mid X_1)$

Product rule general form:

$$P(X_1, \ldots, X_n) = P(x_1 \ldots x_t, x_{t+1} \ldots x_n)$$

$$= P(X_1,\ldots,X_t) P(X_{t+1} \ldots X_n \mid X_1,\ldots,X_t)$$

Chain Rule

Product rule general form:

$$P(X_1, \ldots, X_n) =$$

$$= P(X_1,\ldots,X_t)\, P(X_{t+1} \ldots X_n \mid X_1,\ldots,X_t)$$

Chain rule is derived by successive application of product rule:

t = n - 1

$$P(X_1, \ldots X_{n-1}, X_n) =$$

$$= P(X_1,\ldots,X_{n-1})\, P(X_n \mid X_1,\ldots,X_{n-1})$$

$$= P(X_1,\ldots,X_{n-2})\, P(X_{n-1} \mid X_1,\ldots,X_{n-2})\, P(X_n \mid X_1,\ldots,X_{n-1}) = \ldots$$

$$= P(X_1)\, P(X_2 \mid X_1) \ldots P(X_{n-1} \mid X_1,\ldots,X_{n-2})\, P(X_n \mid X_1,..,X_{n-1})$$

$$= \prod_{i=1}^{n} P(X_i \mid X_1, \ldots, X_{i-1})$$

Chain Rule: Example

P(*cavity* , *toothache*, *catch*) =

P(cavity) * P(toothache | cavity) *

* P(catch | cavity, toothache)

P(*toothache*, *catch*, *cavity*) =

P(toothache) * P(catch | toothache) * P(cavity | toothache, catch)

these and the other four decompositions are OK

Lecture Overview

- Finish Logics
  - Recap Top Down + TD as Search
  - Datalog

Start Stochastic Environments

- Intro to Probability
  - Semantics of Probability
  - Marginalization
  - Conditional Probability and Chain Rule
  - **Bayes' Rule and Independence**

Using conditional probability

- Often you have causal knowledge (forward from cause to evidence):
  - 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 (backwards from evidence to cause):
  - 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 :

$$P(h \mid e) = \frac{P(h \wedge e)}{P(e)} \qquad P(e \mid h) = \frac{P(e \wedge h)}{P(h)}$$

- We can rearrange terms to write

$$P(h \wedge e) = P(h \mid e) \times P(e) \qquad (1)$$

$$P(e \wedge h) = P(e \mid h) \times P(h) \qquad (2)$$

- But

$$P(h \wedge e) = P(e \wedge h) \qquad (3)$$

- From (1) (2) and (3) we can derive

Bayes Rule

$$P(h \mid e) = \frac{P(e \mid h) P(h)}{P(e)} \qquad (3)$$

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%!

Do you always need to revise your beliefs?

…… when your knowledge of **Y**'s value doesn't affect your belief in the value of **X**

DEF. Random variable **X** is marginal independent of random variable **Y** if, for all $x_i \in dom(X)$, $y_k \in dom(Y)$,

$$P(X = x_i \mid Y = y_k) = P(X = x_i)$$

Consequence:

$$P(X = x_i, Y = y_k) = P(X = x_i \mid Y = y_k) \, P(Y = y_k) =$$

$$= P(X = x_i) \, P(Y = y_k)$$

Marginal Independence: Example

X and *Y* are independent iff: $P(X) = P(X|Y) = \frac{P(X,Y)}{P(Y)}$

P(*X*/*Y*) = **P**(*X*) or **P**(*Y*/*X*) = **P**(*Y*) or **P**(X, Y) = **P**(*X*) **P**(*Y*)

That is new evidence Y(or X) does not affect current belief in X (or Y)

|dom| = 4 → Sunny, Cloudy, rainy, snowy

Ex: **P**(*Toothache, Catch, Cavity, Weather*)
= **P**(*Toothache, Catch, Cavity*) P(weather)

JPD requiring 32 entries is reduced to two smaller ones (8 and 4)

JPD: Joint prob. distribution

Learning Goals for today's class

You can:

Given a joint, compute distributions over any subset of the variables

Prove the formula to compute $P(h|e)$

Derive the **Chain Rule** and **the Bayes Rule**

Define **Marginal Independence**

Midterm review

Average 77 ☺

Best 105

Four < 50%

How to learn more from midterm

- Carefully examine your mistakes (and our feedback)
- If you still do not see the correct answer/solution go back to your notes, the slides and the textbook
- If you are still confused come to office hours with specific questions

Next Class

- Conditional Independence
- Belief Networks…….

Assignments

- I will post Assignment 3 this evening
- Assignment2
  - If any of the TAs' feedback is unclear go to office hours
  - If you have questions on the programming part, office hours next Tue (Ken)

Plan for this week

Probability is a rigorous formalism for uncertain knowledge

Joint probability distribution specifies probability of every possible world

Probabilistic queries can be answered by summing over possible worlds

For nontrivial domains, we must find a way to reduce the joint distribution size

Independence *(rare)* and **conditional independence** *(frequent)* provide the tools

Conditional probability (irrelevant evidence)

New evidence may be irrelevant, allowing simplification, e.g.,

- P(*cavity* | *toothache, sunny*) = P(*cavity* | *toothache*)
- We say that Cavity is conditionally independent from Weather (more on this next class)

This kind of inference, sanctioned by domain knowledge, is crucial in probabilistic inference

Bottom-up vs. Top-down

- **Key Idea of top-down:** search backward from a query g to determine if it can be derived from *KB*.

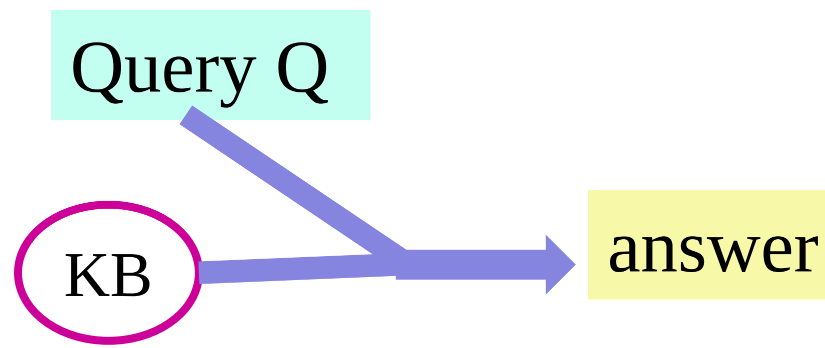
Bottom-up

g is proved if g $\in$ C

When does BU look at the query q?

- Never
- It derives the same q regardless of the query

Top-down



TD performs a backward search starting at q

Inference as Standard Search

- Constraint Satisfaction (Problems):
  - State: assignments of values to a subset of the variables
  - Successor function: assign values to a "free" variable
  - Goal test: set of constraints
  - Solution: possible world that satisfies the constraints
  - Heuristic function: none (all solutions at the same distance from start)
- Planning :
  - State: full assignment of values to features
  - Successor function: states reachable by applying valid actions
  - Goal test: partial assignment of values to features
  - Solution: a sequence of actions
  - Heuristic function: relaxed problem! E.g. "ignore delete lists"
- Query (Top-down/SLD resolution)
  - State: answer clause of the form $yes \leftarrow a_1 \wedge \ldots \wedge a_k$
  - Successor function: all states resulting from substituting first atom a_1 with $b_1 \wedge \ldots \wedge b_m$ if there is a clause $a_1 \leftarrow b_1 \wedge \ldots \wedge b_m$
  - Goal test: is the answer clause empty (i.e. $yes \leftarrow$) ?
  - Solution: the proof, i.e. the sequence of SLD resolutions
  - Heuristic function: e.g. number of atoms in a given answer close

Sound and Complete?

- When you have derived an answer, you can read a bottom up proof in the opposite direction.
- Every top-down derivation corresponds to a bottom up proof and every bottom up proof has a top-down derivation.
- We used this equivalence to prove the soundness and completeness of the SLD proof procedure.

Lecture Overview

- Recap of Lecture 26
- ➡ DataLog
- Logic Wrap up
- Intro to Reasoning Under Uncertainty (time permitting)
  - Motivation
  - Introduction to Probability

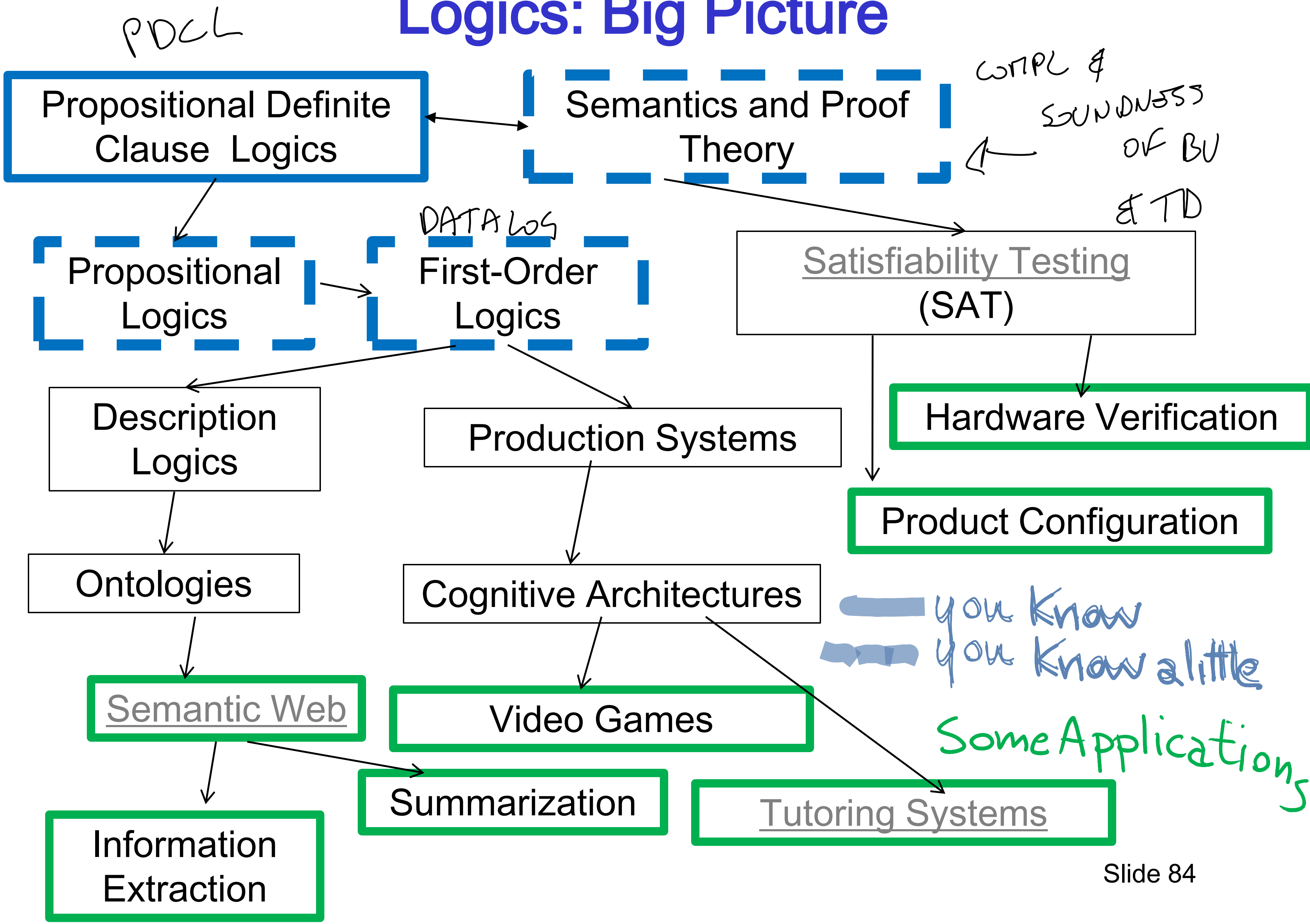
Learning Goals For Logic

- PDCL syntax & semantics
  - Verify whether a logical statement belongs to the language of propositional definite clauses
  - Verify whether an interpretation is a model of a PDCL KB.
  - Verify when a conjunction of atoms is a logical consequence of a KB
- Bottom-up proof procedure
  - Define/read/write/trace/debug the Bottom Up (**BU**) proof procedure
  - Prove that the BU proof procedure is sound and complete
- Top-down proof procedure
  - Define/read/write/trace/debug the Top-down (**SLD**) proof procedure
  - Define it as a search problem
- Datalog
  - Represent simple domains in Datalog
  - Apply the Top-down proof procedure in Datalog

Lecture Overview

- Recap of Lecture 26
- DataLog
- Logic Wrap up
- Intro to Reasoning Under Uncertainty (time permitting)
  - Motivation
  - Introduction to Probability

Logics: Big Picture



Logics: Big picture

- We only covered rather simple logics
  - There are much more powerful representation and reasoning systems based on logics e.g.
    - ✓ full first order logic (with negation, disjunction and function symbols)
    - ✓ second-order logics
    - ✓ non-monotonic logics, modal logics, …
- There are many important applications of logic
  - For example, software agents roaming the web on our behalf
    - ✓ Based on a more structured representation: the semantic web
    - ✓ This is just one example for how logics are used

Semantic Web: Extracting data

- Examples for typical queries
  - How much is a typical flight to Mexico for a given date?
  - What's the cheapest vacation package to some place in the Caribbean in a given week?
    - ✓ Plus, the hotel should have a white sandy beach and scuba diving
- If webpages are based on basic HTML
  - Humans need to scout for the information and integrate it
  - Computers are not reliable enough (yet?)
    - ✓ Natural language processing (NLP) can be powerful (see Watson and Siri!)
    - ✓ But some information may be in pictures (beach), or implicit in the text, so existing NLP techniques still don't get all the info

More structured representation: the Semantic Web

- Beyond HTML pages only made for humans
- Languages and formalisms based on *description* logics that allow websites to include rich, explicit information on
  - relevant concepts, individual and their relationships \
  - Goal: software agents that can roam the web and carry out sophisticated tasks on our behalf, based on these richer representations
- Different than searching content for keywords and popularity.
  - Infer meaning from content based on metadata and assertions that have already been made.
  - Automatically classify and integrate information
- For further material, P&M text, Chapter 13. Also
  - the Introduction to the Semantic Web tutorial given at 2011 Semantic TechnologyConference http://www.w3.org/People/Ivan/CorePresentations/SWTutorial/

Examples of ontologies for the Semantic Web

"Ontology": logic-based representation of the world

- eClassOwl: eBusiness ontology
  - for products and services
  - 75,000 classes (types of individuals) and 5,500 properties
- National Cancer Institute's ontology: 58,000 classes
- Open Biomedical Ontologies Foundry: several ontologies
  - including the Gene Ontology to describe
    - ✓ gene and gene product attributes in any organism or protein sequence
- OpenCyc project: a 150,000-concept ontology including
  - Top-level ontology
    - ✓ describes general concepts such as numbers, time, space, etc
  - Hierarchical composition: superclasses and subclasses
  - Many specific concepts such as "OLED display", "iPhone"

A different example of applications of logic

Cognitive Tutors (http://pact.cs.cmu.edu/)

- computer tutors for a variety of domains (math, geometry, programming, etc.)
  - Provide individualized support to problem solving exercises, as good human tutors do
  - Rely on logic-based, detailed computational models of skills and misconceptions underlying a learning domain.
- CanergieLearning (http://www.carnegielearning.com/):
  - a company that commercializes these tutors, sold to hundreds of thousands of high schools in the USA

Inference as Standard Search

- Constraint Satisfaction (Problems):
  - State: assignments of values to a subset of the variables
  - Successor function: assign values to a "free" variable
  - Goal test: set of constraints
  - Solution: possible world that satisfies the constraints
  - Heuristic function: none (all solutions at the same distance from start)
- Planning :
  - State: full assignment of values to features
  - Successor function: states reachable by applying valid actions
  - Goal test: partial assignment of values to features
  - Solution: a sequence of actions
  - Heuristic function: relaxed problem! E.g. "ignore delete lists"
- Query (Top-down/SLD resolution)
  - State: answer clause of the form $yes \leftarrow a_1 \wedge \dots \wedge a_k$
  - Successor function: all states resulting from substituting first atom a_1 with $b_1 \wedge \dots \wedge b_m$ if there is a clause $a_1 \leftarrow b_1 \wedge \dots \wedge b_m$
  - Goal test: is the answer clause empty (i.e. $yes \leftarrow$) ?
  - Solution: the proof, i.e. the sequence of SLD resolutions
  - Heuristic function: ?????