Reasoning Under Uncertainty: Introduction to Probability

CPSC 322 – Uncertainty 1

Textbook §6.1

March 16, 2011

Coloured Cards

- If you lost/forgot your set,
  please come to the front and pick up a new one
  - We'll use them quite a bit in the uncertainty module

Lecture Overview

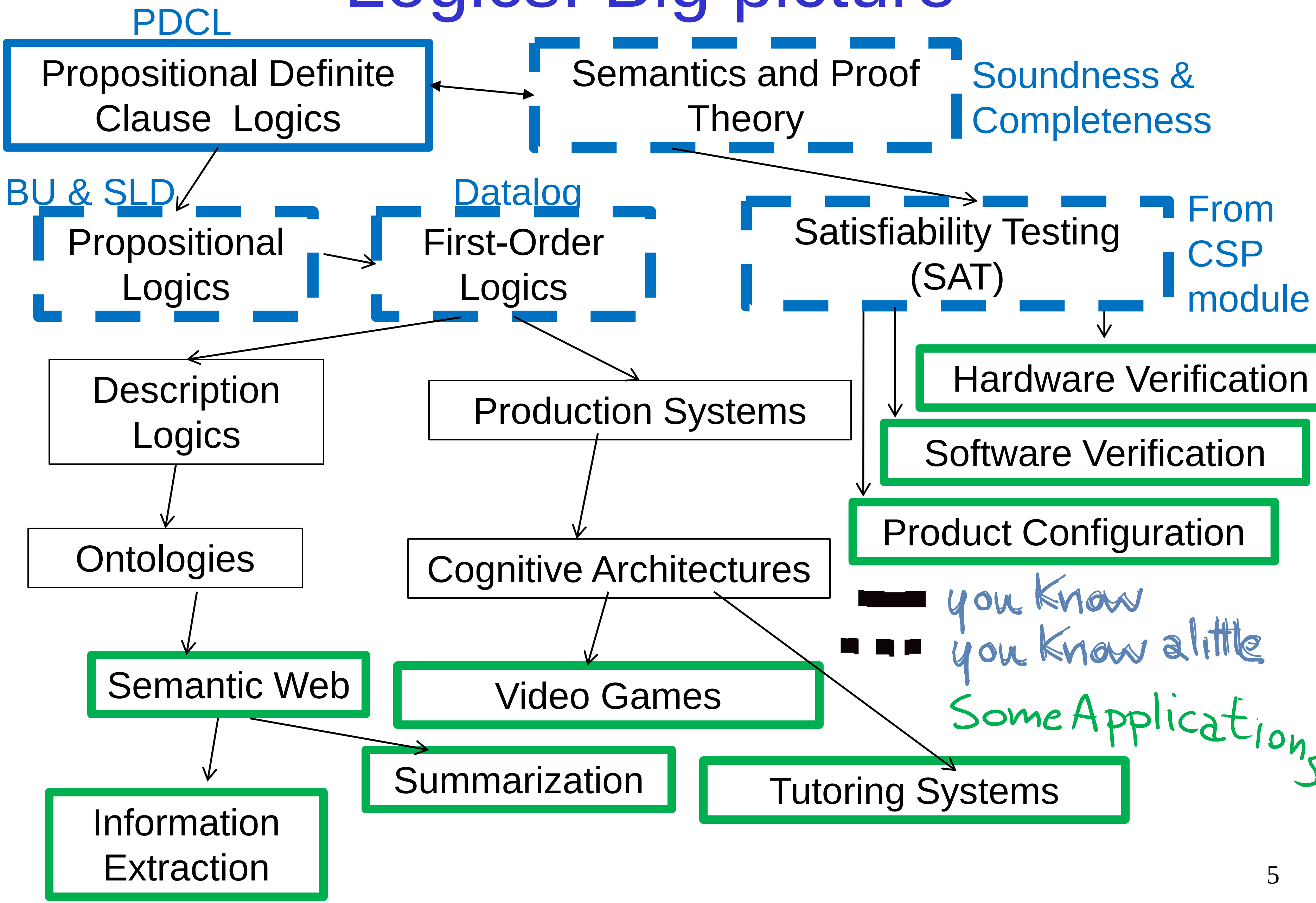
- Logics wrap-up: big picture

- Reasoning Under Uncertainty
  - Motivation
  - Introduction to Probability
    - Random Variables and Possible World Semantics
    - Probability Distributions and Marginalization
    - Time-permitting: Conditioning

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 (as a search problem)
- Datalog
  - Represent simple domains in Datalog
  - Apply the Top-down proof procedure in Datalog

Logics: Big picture



Logics: Big picture

- We only covered rather simple logics
  - There are much more powerful representation and reasoning systems based on logics
- Logics have many applications
  - See previous slide
  - Let's see the 2-slide version of one example: the Semantic Web

Example application of logics: the Semantic Web

- Beyond HTML pages only made for humans
- Languages and formalisms based on logics that allow websites to include information in a more structured format
  - Goal: software agents that can roam the web and carry out sophisticated tasks on our behalf
  - This is very different than searching content for keywords and popularity!

- For further references, see, e.g. tutorial given at 2009 Semantic Technology Conference:

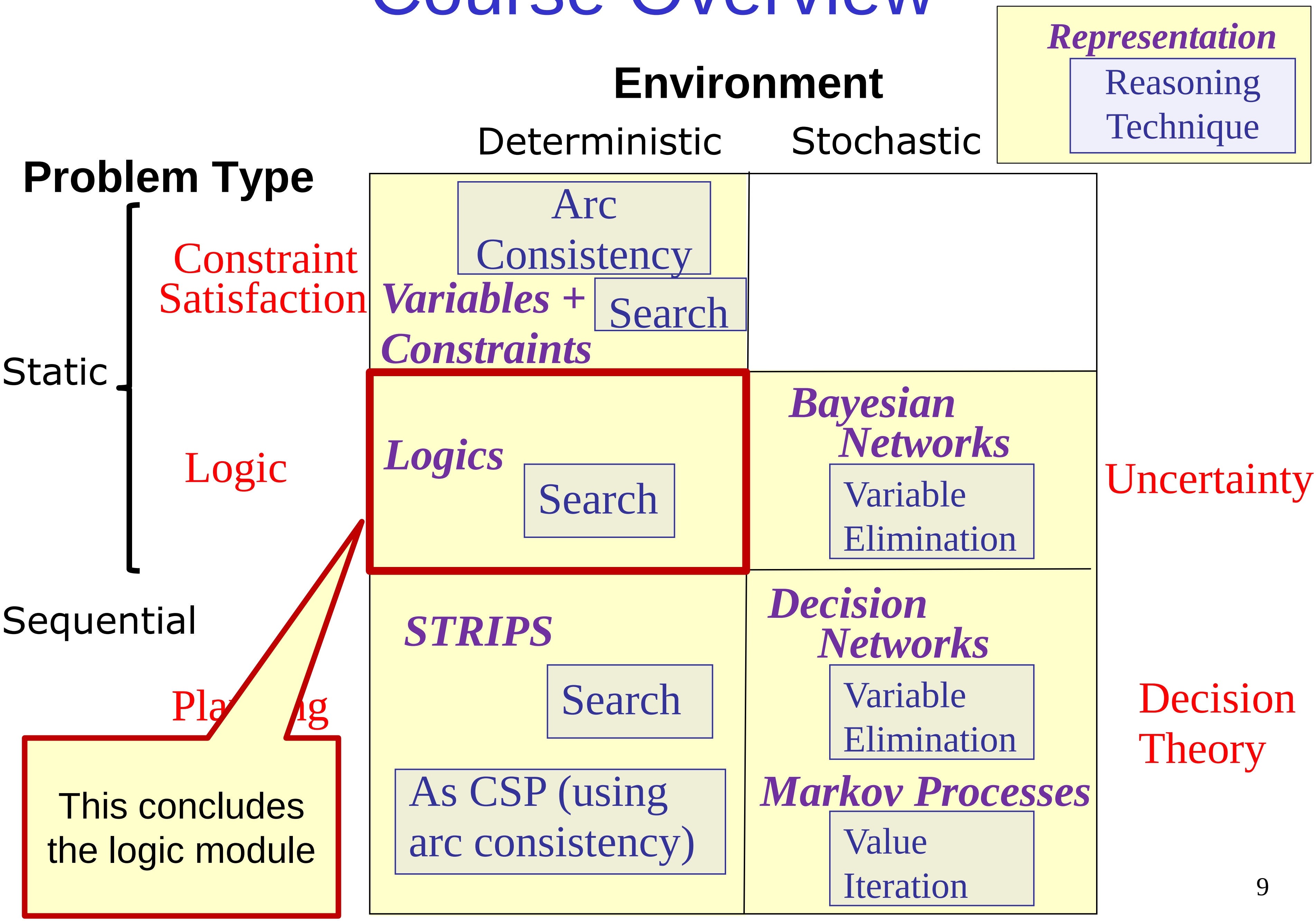
  http://www.w3.org/2009/Talks/0615-SanJose-tutorial-IH

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
    - annotation terminology and data
- 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"

Course Overview



Course Overview

Course Module

Representation

Reasoning Technique

Environment

Deterministic | Stochastic

Problem Type

Static

Constraint Satisfaction

Variables + Constraints

Arc Consistency

Search

Logic

Logics

Search

Bayesian Networks

Variable Elimination

Uncertainty

For the rest of the course, we will consider uncertainty

Sequential

Planning

STRIPS

Search

As CSP (using arc consistency)

Decision Networks

Variable Elimination

Markov Processes

Value Iteration

Decision Theory

Lecture Overview

- Logics wrap-up: big picture
- Reasoning Under Uncertainty
  - Motivation
  - Introduction to Probability
    - Random Variables and Possible World Semantics
    - Probability Distributions and Marginalization
    - Time-permitting: Conditioning

Types of uncertainty (from Lecture 2)

- Sensing Uncertainty:
  - The agent cannot fully observe a state of interest
  - E.g.: Right now, how many people are in this room? In this building?
  - E.g.: What disease does this patient have?
- Effect Uncertainty:
  - The agent cannot be certain about the effects of its actions
  - E.g.: If I work hard, will I get an A?
  - E.g.: Will this drug work for this patient?

Motivation for uncertainty

- To act in the real world, we almost always have to handle uncertainty (both effect and sensing uncertainty)
  - Deterministic domains are an abstraction
    - Sometimes this abstraction enables more powerful inference
  - Now we don't make this abstraction anymore
    - Our representation becomes more expressive and general
- AI's focus shifted from logic to probability in the 1980s
  - The language of probability is very expressive and general
  - New representations enable efficient reasoning
    - We will see some of these, in particular Bayesian networks
  - Reasoning under uncertainty is the "new" AI
  - See, e.g., Faculty Lecture Series talk tomorrow:
    - "The Cancer Genome and Probabilistic Models" DMP 110, 3:30-4:50

Interesting article about AI and uncertainty

- "The machine age"
  - by Peter Norvig (head of research at Google)
  - New York Post, 12 February 2011
  - http://www.nypost.com/f/print/news/opinion/opedcolumnists/the_machine_age_tM7xPAv4pI4JslK0M1JtxI

  - "The things we thought were hard turned out to be easier."
    - Playing grandmaster level chess,
      or proving theorems in integral calculus
  - "Tasks that we at first thought were easy turned out to be hard."
    - A toddler (or a dog) can distinguish hundreds of objects (ball, bottle, blanket, mother, etc.) just by glancing at them
    - Very difficult for computer vision to perform at this level
  - "Dealing with uncertainty turned out to be more important than thinking with logical precision."
    - AI's focus shifted from Logic to Probability (in the late 1980s)
    - Reasoning under uncertainty (and lots of data) are key to progress

Lecture Overview

- Logics wrap-up: big picture
- Reasoning Under Uncertainty
  - Motivation
  - Introduction to Probability
    - Random Variables and Possible World Semantics
    - Probability Distributions and Marginalization
    - Time-permitting: Conditioning

Probability as a formal measure of uncertainty/ignorance

- Probability measures an agent's degree of belief on events
  - It does not measure how true an event is
  - Events are true or false. We simply might not know exactly which one
  - Example:
    - I roll a fair die. What is the probability that the result is a “6”?

Probability as a formal measure of uncertainty/ignorance

- Probability measures an agent's degree of belief on events
  - It does not measure how true an event is
  - Events are true or false. We simply might not know exactly which one
  - Example:
    - I roll a fair die. What is the probability that the result is a “6”?
      - It is 1/6 ≈ 16.7%.
      - The result is either a “6” or not. But we don’t know which one.
    - I now look at the die. What is the probability now?
      - Your probability hasn’t changed: 1/6 ≈ 16.7%
      - My probability is either 1 or 0 (depending on what I observed)
    - What if I tell some of you the result is even?
      - Their probability increases to 1/3 ≈ 33.3%
        (assuming they know I say the truth)
    - Different agents can have different degrees of belief in an event

Probability as a formal measure of uncertainty/ignorance

- Probability measures an agent's degree of belief on events
  - It does not measure how true an event is
  - Events are true or false. We simply might not know exactly which one
    - Different agents can have different degrees of belief in an event

- Belief in a proposition *f* can be measured in terms of a number between 0 and 1 – this is the probability of *f*
  - P("roll of fair die came out as a 6") = 1/6 ≈ 16.7% = 0.167
  - Using probabilities between 0 and 1 is purely a convention.

- P(f) = 0 means that f is believed to be

Probably true	Probably false	Definitely false	Definitely true

Probability as a formal measure of uncertainty/ignorance

- Probability measures an agent's degree of belief on events
  - It does not measure how true an event is
  - Events are true or false. We simply might not know exactly which one
    - Different agents can have different degrees of belief in an event

- Belief in a proposition *f* can be measured in terms of a number between 0 and 1 – this is the probability of *f*
  - P("roll of fair die came out as a 6") = 1/6 ≈ 16.7% = 0.167
  - Using probabilities between 0 and 1 is purely a convention.

- P(f) = 0 means that *f* is believed to be
  - Definitely false: the probability of f being true is zero.
- Likewise, P(f) = 1 means *f* is believed to be definitely true

Probability Theory and Random Variables

- Probability Theory: system of axioms and formal operations for sound reasoning under uncertainty

- Basic element: random variable X
  - X is a variable like the ones we have seen in CSP/Planning/Logic, but the agent can be uncertain about the value of X
  - As usual, the domain of a random variable *X*, written dom(X), is the set of values *X* can take

- Types of variables
  - Boolean: e.g., *Cancer* (does the patient have cancer or not?)
  - Categorical: e.g., *CancerType* could be one of <*breastCancer, lungCancer, skinMelanomas*>
  - Numeric: e.g., Temperature

  - We will focus on Boolean and categorical variables

Possible Worlds Semantics

- A possible world w specifies an assignment to each random variable

- Example: we model only 2 Boolean variables *Smoking* and *Cancer*, how many distinct possible worlds are there?

Possible Worlds Semantics

- A possible world w specifies an assignment to each random variable

- Example: we model only 2 Boolean variables *Smoking* and *Cancer*. Then there are 2^2=4 distinct possible worlds:

w_1: *Smoking = T* $\wedge$ *Cancer = T*
w_2: *Smoking = T* $\wedge$ *Cancer = F*
w_3: *Smoking = F* $\wedge$ *Cancer = T*
w_4: *Smoking = T* $\wedge$ *Cancer = T*

Smoking	*Cancer*
T	T
T	F
F	T
F	F

- $w \models X{=}x$ means variable X is assigned value x in world w
- Define a nonnegative measure $\mu(w)$ to possible worlds w such that the measures of the possible worlds sum to 1

-The probability of proposition f is defined by: $$p(f) = \sum_{w \models f} \mu(w)$$

Possible Worlds Semantics

- New example: weather in Vancouver
  - Modeled as one Boolean variable:
    - *Weather* with domain {sunny, cloudy}
  - Possible worlds:
    w_1: *Weather = sunny*
    w_2: *Weather = cloudy*

Weather	*p*
sunny	0.4
cloudy	

- Let's say the probability of sunny weather is 0.4
  - *I.e. p(Weather = sunny)* = 0.4
  - What is the probability of *p(Weather = cloudy)*?

We don't have enough information to compute that probability

0.4

1

0.6

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

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

Possible Worlds Semantics

- New example: weather in Vancouver
  - Modeled as one Boolean variable:
    - *Weather* with domain {sunny, cloudy}
  - Possible worlds:
    w_1: *Weather = sunny*
    w_2: *Weather = cloudy*

Weather	*p*
sunny	0.4
cloudy	**0.6**

- Let's say the probability of sunny weather is 0.4
  - *I.e. p(Weather = sunny)* = 0.4
  - What is the probability of *p(Weather = cloudy)*?
    - p(*Weather = sunny*) = 0.4 means that $\mu(w_1)$ is 0.4
    - $\mu(w_1)$ and $\mu(w_2)$ have to sum to 1 (those are the only 2 possible worlds)
    - So $\mu(w_2)$ has to be 0.6, and thus p(*Weather = cloudy*) = 0.6

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

One more example

- Now we have an additional variable:
  - Temperature, modeled as a categorical variable with domain {hot, mild, cold}
  - There are now 6 possible worlds:

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	?

  - What's the probability of it being cloudy and cold?

0.1 0.2 0.3 1

- Hint: 0.10 + 0.20 + 0.10 + 0.05 + 0.35 = 0.8

One more example

- Now we have an additional variable:
  - Temperature, modeled as a categorical variable with domain {hot, mild, cold}
  - There are now 6 possible worlds:

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	**0.2**

  - What's the probability of it being cloudy and cold?
    - It is 0.2: the probability has to sum to 1 over all possible worlds

Lecture Overview

- Logics wrap-up: big picture
- Reasoning Under Uncertainty
  - Motivation
  - Introduction to Probability
    - Random Variables and Possible World Semantics
    - Probability Distributions and Marginalization
    - Time-permitting: Conditioning

Probability Distributions

Consider the case where possible worlds are simply assignments to one random variable.

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

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

- When dom(X) is infinite we need a probability density function
- We will focus on the finite case

Joint Distribution

- The joint distribution over random variables $X_1, \ldots, X_n$:
  - a probability distribution over the joint random variable $\langle X_1, \ldots, X_n \rangle$ with domain $dom(X_1) \times \ldots \times dom(X_n)$ (the Cartesian product)

- Example from before
  - Joint probability distribution over random variables Weather and Temperature
  - Each row corresponds to an assignment of values to these variables, and the probability of this joint assignment

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

  - In general, each row corresponds to an assignment $X_1= x_1, \ldots, X_n= x_n$ and its probability $P(X_1= x_1, \ldots ,X_n= x_n)$
  - We also write $P(X_1= x_1 \wedge \ldots \wedge X_n= x_n)$
  - The sum of probabilities across the whole table is 1.

Marginalization

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

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

  – We also write this as $P(X) = \Sigma_{z \in dom(Z)} P(X, Z = z)$.

- 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	?
mild	?
cold	?

Marginalization

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

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

  – We also write this as $P(X) = \Sigma_{z \in dom(Z)}\ P(X, Z = z)$.

- 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	??
mild	
cold	

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

Marginalization

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

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

  – We also write this as $P(X) = \Sigma_{z \in \mathrm{dom}(Z)}\ P(X, Z = z)$.

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

Marginalization

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

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

  – We also write this as $P(X) = \Sigma_{z \in dom(Z)}\ P(X, Z = z)$.

- 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	

0.20 0.35 0.85 0.55

Marginalization

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

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

  – We also write this as $P(X) = \Sigma_{z \in dom(Z)}\ P(X, Z = z)$.

- 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	??

0.70 0.30 0.20 0.10

Marginalization

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

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

  – We also write this as $P(X) = \Sigma_{z \in dom(Z)}\ P(X, Z = z)$.

- 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.

Marginalization

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

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

  - We also write this as $P(X) = \Sigma_{z \in dom(Z)}\ P(X, Z = z)$.

- You can marginalize out any of the variables

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

Weather	*µ(w)*
sunny	0.40
cloudy	

P(Weather=sunny) =
P(Weather=sunny, Temperature = hot)
+ P(Weather=sunny, Temperature = mild)
+ P(Weather=sunny, Temperature = cold)
= 0.10 + 0.20 + 0.10 = 0.40

Marginalization

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

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

  – We also write this as $P(X) = \Sigma_{z \in dom(Z)}\ P(X, Z = z)$.

- You can marginalize out any of the variables

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

Weather	*µ(w)*
sunny	0.40
cloudy	0.60

Marginalization

- We can also marginalize out 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

Weather	*µ(w)*
sunny	0.40
cloudy	

Marginalizing out variables Wind and Temperature, i.e. those are the ones being removed from the distribution

Marginalization

- We can also get marginals for more than one variable

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

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

Weather	*Temperature*	*µ(w)*
sunny	hot	0.10
sunny	mild	
sunny	cold	
cloudy	hot	
cloudy	mild	
cloudy	cold	

Learning Goals For Today's Class

- 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 (JPD)
- Given a JPD
  - Marginalize over specific variables
  - Compute distributions over any subset of the variables

- Heads up: study these concepts, especially marginalization
  - If you don't understand them well you will get lost quickly

Lecture Overview

- Logics wrap-up: big picture
- Reasoning Under Uncertainty
  - Motivation
  - Introduction to Probability
    - Random Variables and Possible World Semantics
    - Probability Distributions and Marginalization
    - Time-permitting: Conditioning

Conditioning

- Conditioning species how to revise beliefs based on new information.
- You build a probabilistic model 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.

Example for conditioning

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

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

- Now, you look outside and see that it's sunny
  - Your knowledge of the weather affects your degree of belief in the temperature
  - The conditional probability distribution for temperature given that it's sunny is:
  - We will see how to compute this.

T	*P(T|W=sunny)*
hot	0.25
mild	0.50
cold	0.25

Semantics of Conditioning

- Evidence e rules out possible worlds incompatible with e.
- We can represent this using a new measure, μ_e, over possible worlds

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

Definition (conditional probability)
The conditional probability of formula h given evidence e is

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

Example for conditioning

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

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

Weather	*μ(w)*
sunny	0.40
cloudy	0.60

- What is P(Temperature| Weather = sunny)?
- We know $P(h|e) = \frac{P(h \wedge e)}{P(e)}$
  – h = “Temperature=t”, e= “Weather = sunny”

Example for conditioning

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

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

Weather	*μ(w)*
sunny	0.40
cloudy	0.60

- What is P(Temperature| Weather = sunny)?
- We know $P(h|e) = \frac{P(h \wedge e)}{P(e)}$
  - h = “Temperature=t”, e= “Weather = sunny”
  - P(Temperature=t ∧ Weather = sunny)
  - P(Weather = sunny)

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