Reasoning Under Uncertainty: Introduction to Probability

Alan Mackworth

UBC CS 322 - Uncertainty 1

March 11, 2013

Textbook §6.1, 6.1.1, 6.1.3

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

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

Course Overview

Course Module: *Representation* — Reasoning Technique

Environment

Problem Type	Course Module	Deterministic	Stochastic	Course Module
Static	Constraint Satisfaction	*Variables + Constraints* — Search; Arc Consistency		
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

Lecture Overview

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

Two main sources of uncertainty

(From Lecture 2)

- Sensing Uncertainty: The agent cannot fully observe a state of interest.

  For example:
  - Right now, how many people are in this room? In this building?
  - What disease does this patient have?
  - Where is the soccer player behind me?

- Effect Uncertainty: The agent cannot be certain about the effects of its actions.

  For example:
  - If I work hard, will I get an A?
  - Will this drug work for this patient?
  - Where will the ball go when I kick it?

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 main 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 part of the 'new' AI
  - This is not a dichotomy: framework for probability is logical!
  - New frontier: combine logic and probability

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_tM7xPAv4pI4JsIK0M1JtxI
  - "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, …) 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."
    - Reasoning under uncertainty (and lots of data) are key to progress

Lecture Overview

- 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 in propositions about states of the world
  - *It does not measure how true a proposition is.*
  - Propositions are true or false. We simply may not know exactly which.
  - Example:
    - I roll a fair dice. 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 in truth of propositions about states of the world
  - It does not measure how true a proposition is
  - Propositions are true or false. We simply may not know exactly which.
  - Example:
    - I roll a fair dice. What is ‘the’ (my) probability that the result is a ‘6’?
      - It is $1/6 \approx 16.7\%$.
      - The result is either a ‘6’ or not. But I don’t know which one.
    - I now look at the dice. What is ‘the’ (my) probability now?
      - *Your* probability hasn’t changed: $1/6 \approx 16.7\%$
      - *My* probability is now either 1 or 0, depending on what I observed.
    - What if I tell some of you the result is even?
      - *Their* probability increases to $1/3 \approx 33.3\%$
        (assuming they believe I speak the truth)
    - Different agents can have different degrees of belief in (probabilities for) a proposition conditioned on the evidence they have.

Probability as a formal measure of uncertainty/ignorance

- Probability measures an agent's degree of belief in truth of propositions about states of the world
- It does not measure how true a proposition is
  - Propositions are true or false.
  - Different agents can have different degrees of belief in the truth of a proposition
  - This is the **subjective** interpretation of probability.
- 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 in truth of propositions about states of the world
- It does not measure how true a proposition is
  - Propositions are true or false.
  - Different agents can have different degrees of belief in the truth of a proposition
  - This is the **subjective** interpretation of probability.
- 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 logical 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 (integer or real)
  - We will focus on Boolean and categorical variables

Possible Worlds Semantics

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

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

w_1: *Smoking = T* ∧ *Cancer = T*
w_2: *Smoking = T* ∧ *Cancer = F*
w_3: *Smoking = F* ∧ *Cancer = T*
w_4: *Smoking = T* ∧ *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 $\mu(w)$ on possible worlds w such that the measures on 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 categorical 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

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

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

- 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) → [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 $\mathrm{dom}(X_1) \times \ldots \times \mathrm{dom}(X_n)$ (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*	$\mu(w)$
sunny	hot	0.10
sunny	mild	0.20
sunny	cold	0.10
cloudy	hot	0.05
cloudy	mild	0.35
cloudy	cold	0.20

  – 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:

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

– 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:

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

–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	

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:

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

–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:

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

–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:

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

  – 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:

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

  – 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

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

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

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

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

Weather	*Temperature*	*P(W,T)*
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	*P(T)*
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

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*	*$\mu(w)$*
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 *now* 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 T*	*$\mu(w)$ P(W,T)*	*$\mu_e(w)$*
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, μ_e, over possible worlds

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

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, μ_e, over possible worlds

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

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*	*$\mu(w)$ P(W,T)*	*$\mu_e(w)$ P(T|W=sunny)*
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, μ_e, over possible worlds

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

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

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