EECS 496: Sequential Decision Making

Soumya Ray

sray@case.edu

Office: Olin 516

Office hours: T 4-5:30 or by appointment

Recap

- How are satisfiability and inference related?
- What is a clause? What is CNF?
- How do we set up satisfiability as goal directed search? (the 4 things)
- What is the general strategy behind DPLL?
- What is the pure literal heuristic?
- What is unit propagation?
- What is early termination?
- To set up satisfiability as optimization, the objective function is _____, the states are _____ and the operators are ___________.
- How does WalkSAT work?
- What parameters govern the hardness of a SAT problem?
- Why?
- What is a "phase transition" behavior?
- Where are the hardest SAT problems located?

Today

- Probability theory (Ch 13, Russell and Norvig)

Probability Theory

- A language that *augments* propositional logic with "degrees of belief," and associated mechanics for reasoning in this augmented language

I think it is 60% likely that it will rain tomorrow.

$RainTomorrow = true$ 60%

(proposition) (degree of belief)

Random Variable (R.V.)

- A variable that refers to an uncertain fact
  - Analogous to proposition symbol
  - Has a domain that can be discrete or continuous
    - For this class, focus on discrete case

- For each value (or set of values), we can specify a *degree of belief* that shows how much we believe the stated fact---this is the *probability* associated with the fact
  - Denoted $\Pr(.)$

Example

- $RainTomorrow \in$ {True, False}
  - $\Pr(RainTomorrow = True) = 0.6$

- $Current\_X\_Position \in (-\infty, +\infty)$
  - $\Pr(-1 \le Current\_X\_Position) = 0.2$

Atomic Event

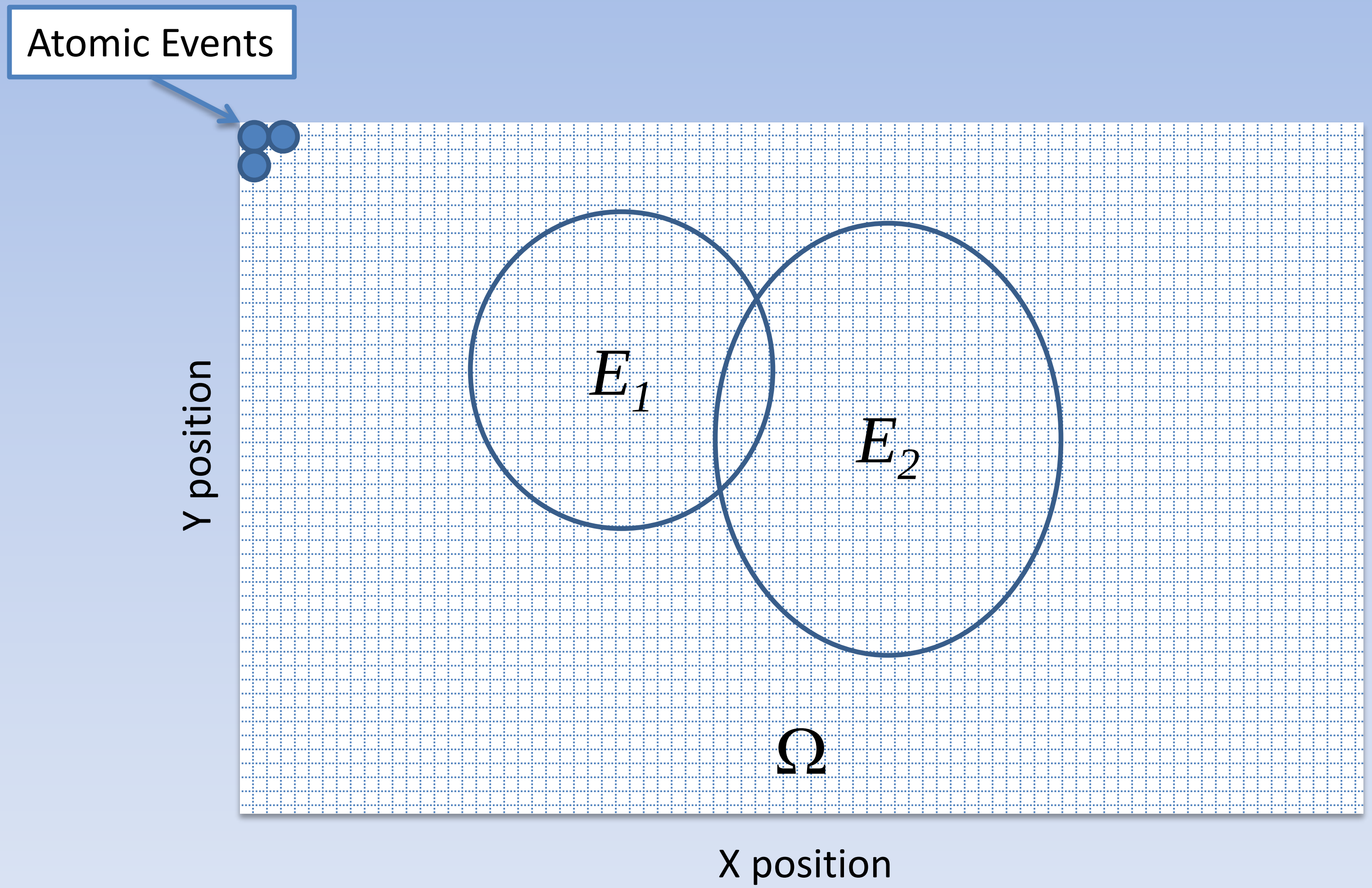
- If the state of the world is described by n r.v.'s and we assign values to all of them, this defines an atomic event
  - (Analog of a row in a truth table)

- Example: suppose a footman is in a grid maze and is uncertain about an enemy archer's (x, y) location. Then $(x=2, y=3)$ could be an atomic event.

Events and the Sample Space

- Atomic events are mutually exclusive and exhaustive
  - At most one can be the true state of affairs
  - The true state of affairs must be one of them
- An "event" is a collection of atomic events
  - Example: the event $\{x=2\}$ is the collection of atomic events $\{(x=2, y=1), (x=2, y=2), (x=2, y=3),\ldots\}$
- The "sample space" is the collection of all possible atomic events (Ω)
  - Analog of the full truth table

Picture



Joint Probability

- Just like we assign degrees of belief to single r.v.'s, we can do the same for groups of r.v.'s
  - $\Pr(RainTomorrow{=}Yes,\ CloudyTomorrow{=}Yes) = 0.99$
  - $\Pr(-1 \leq x, y \leq 1) = 0.2$
  - In particular, we can assign degrees of belief to atomic events

Axioms of Probability

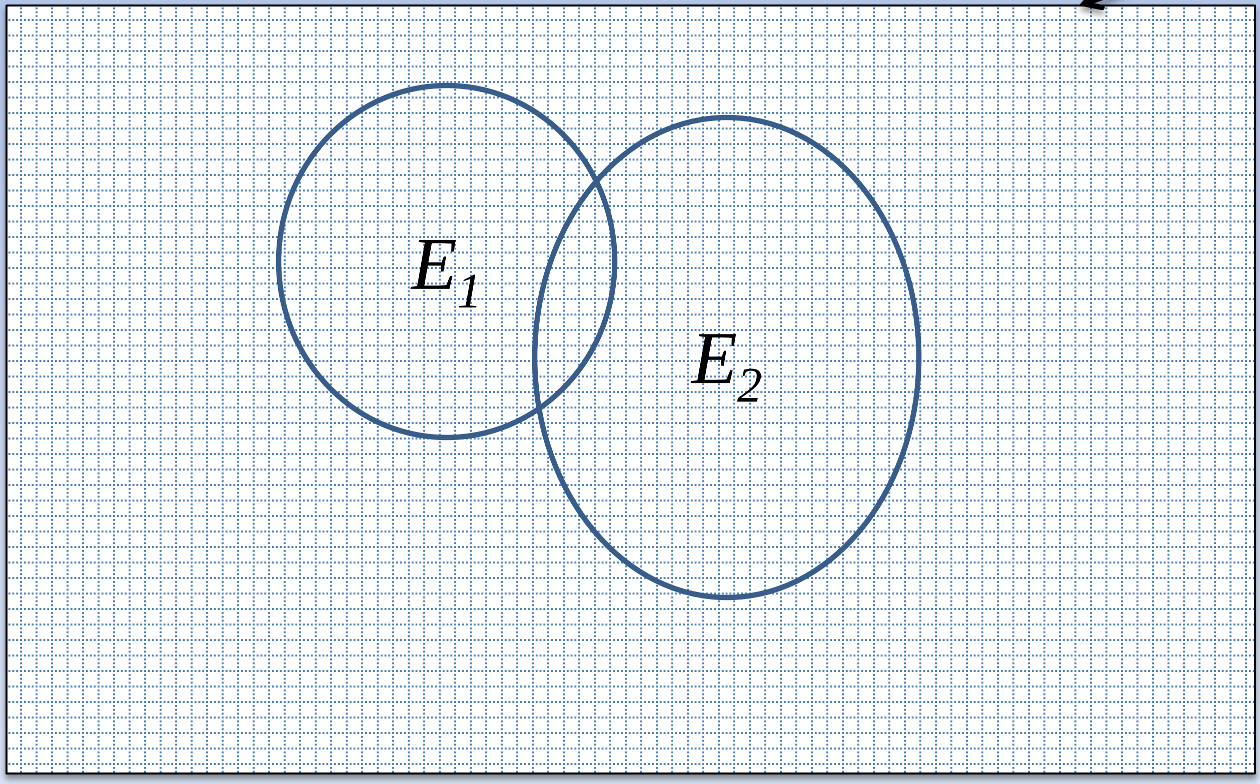
- For any event E, $0 \leq \Pr(E) \leq 1$
- $\Pr(\Omega)=1$
- For *mutually disjoint* events, the probability of the union is given by:

$$\Pr(\bigcup_{i=1} E_i) = \sum_{i=1} \Pr(E_i)$$

In particular this must apply to atomic events.

Picture

Sample Space: Total “area”=1



$Pr(E)$=area under E

Using the axioms

- Various other facts can be deduced from these axioms

- Suppose E is some event and $\bar{E}$ is the event in Ω that includes everything not in E (the "complement" of E). What is $Pr(\bar{E})$?

Rationality and Probability Theory

- Could there be other ways of representing uncertainty?
  - Dempster-Shafer, “Fuzzy” logic, etc

- But probability theory has a major positive result: suppose someone’s degrees of belief for some set of events does NOT obey the axioms of probability. Then there is a way to bet against them such that they will always lose money (utility) over time. (Bruno de Finetti 1931)

Probability Density Functions

- Earlier we defined probabilities associated with r.v.'s: $\Pr(RainTomorrow=Yes)=0.8$
- A function that maps *every* value of an r.v. to a probability is called a probability density function (p.d.f.)

$$p_{RainTomorrow}(x) = \begin{cases} 0.8 \text{ if } x = Yes \\ 0.2 \text{ if } x = No \end{cases}$$

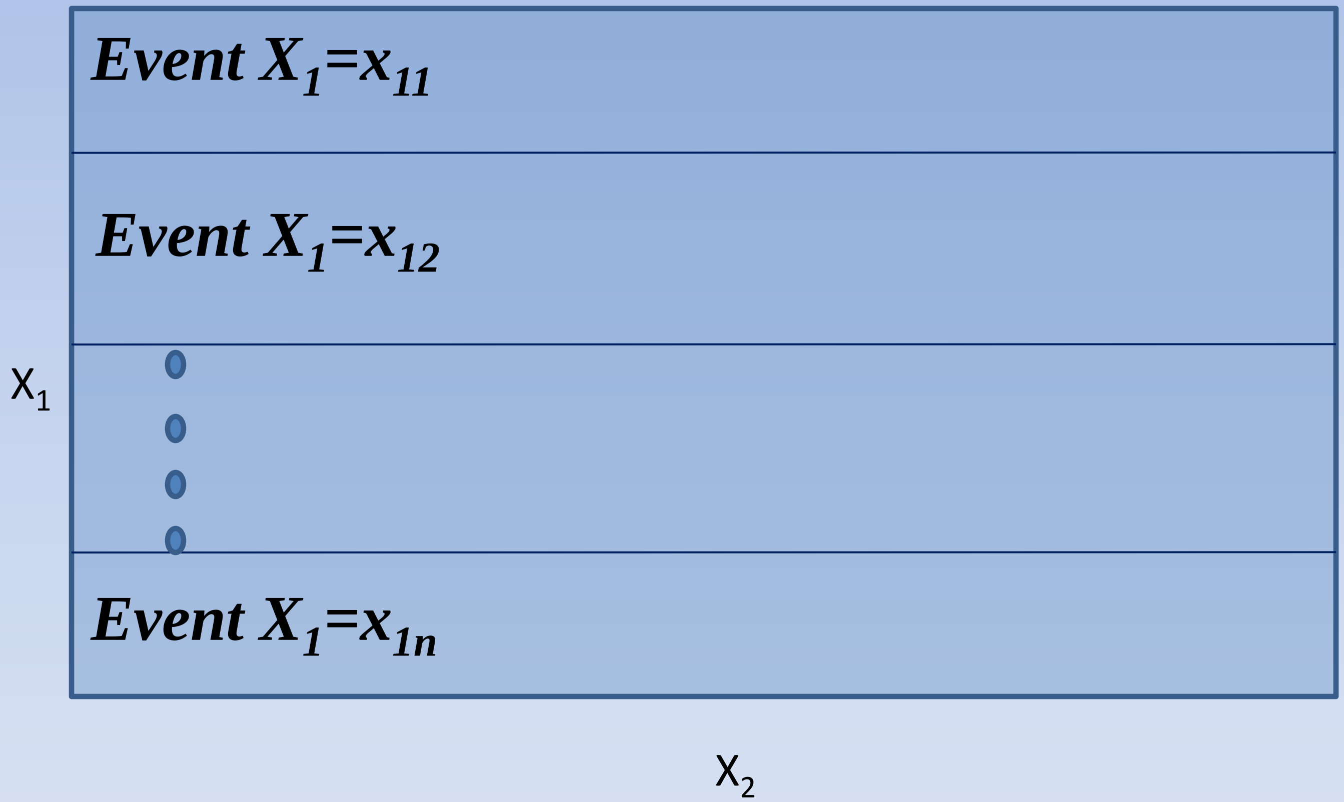
$$p_X(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}},\ x \in \{-\infty, +\infty\}$$

Joint PDF

- Using joint probability, we can define joint density functions for collections of random variables

$$p_{R,C}(R = x, C = y) = \begin{cases} 0.5 \text{ if } x = Yes, y = Yes \\ 0.2 \text{ if } x = No, y = Yes \\ 0.2 \text{ if } x = Yes, y = No \\ 0.1 \text{ if } x = No, y = No \end{cases}$$

All PDFs must sum to 1



Example

Joint Probability Density Function

CloudyTomorrow	RainTomorrow	WetGrass	Probability
No	No	No	0.4
No	No	Yes	0.01
No	Yes	No	0
No	Yes	Yes	0.01
Yes	No	No	0.15
Yes	No	Yes	0.02
Yes	Yes	No	0.01
Yes	Yes	Yes	0.4

Atomic Event

Event

Sample Space

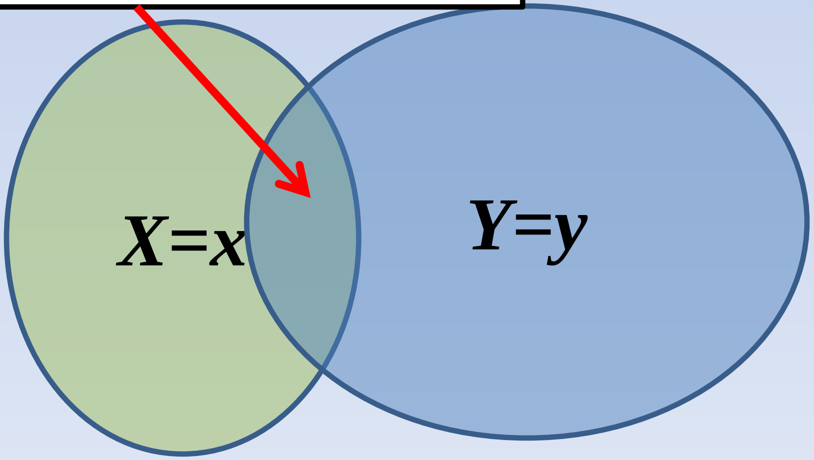
Terminology and Results

Conditional Probability

- The conditional probability of ***X* given *Y*** is:

$$p_{X|Y}(X = x \mid Y = y) = \frac{p_{X,Y}(X = x, Y = y)}{p_Y(Y = y)}$$

***X*=*x*, *Y*=*y* ("," means AND)**



Product Rule

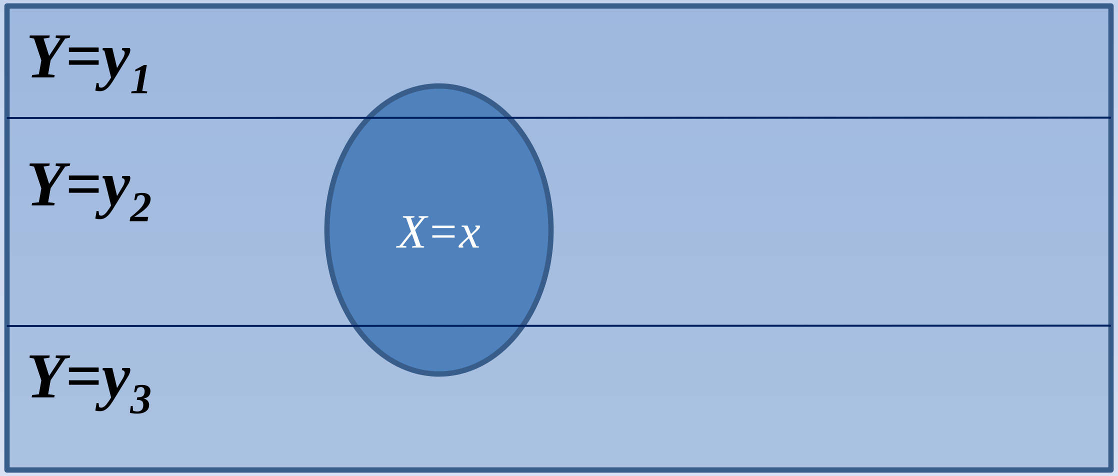
- From the definition of conditional probability:

$$p_{X,Y}(X = x, Y = y) =$$
$$p_Y(Y = y)\, p_{X|Y}(X = x \mid Y = y)$$

Marginalization ("Summing out")

- For any two random variables X and Y:

$$p_X(X=x)=\sum_y p_{X,Y}(X=x,Y=y)$$



Conditioning

$$p(X = x) = \sum_{y} p(X = x, Y = y)$$

$\boxed{\text{Marginalization}}$

$$= \sum_{y} p(X = x \mid Y = y) p(Y = y)$$

$\boxed{\text{Product Rule}}$

Bayes' Rule

(Rev. Thomas Bayes 1763)

$$p(C=c \mid E=e) = \frac{p(C=c, E=e)}{p(E=e)}$$

Def. of Conditional Prob.

$$= \frac{p(E=e \mid C=c)\,p(C=c)}{p(E=e)}$$

Product Rule

$$= \frac{p(E=e \mid C=c)\,p(C=c)}{\sum_{c'} p(E=e \mid C=c')\,p(C=c')}$$

Conditioning

The importance of Bayes' Rule

- Let C be a random variable with values that are possible "causes"
- Let E denote a random variable with values that are possible effects of each cause
- It is often easy to specify $p(E=e|C=c)$, much harder to specify $p(C=c|E=e)$
- Bayes' Rule therefore allows us to **reason backwards over uncertain events**--- fundamental to *learning*

Example

- Lung cancer can be caused by smoking or by a genetic defect. 5% of the population are smokers. 2 in 3 who smoke and 1 in 100 who don’t get the disease.
- Suppose X has lung cancer. What is the probability X smokes?

Example

$$P(S) = 0.05, P(LC \mid S) = 0.67, P(LC \mid \overline{S}) = 0.01$$

$$P(S \mid LC) = \frac{P(LC \mid S)P(S)}{P(LC \mid S)P(S) + P(LC \mid \overline{S})P(\overline{S})}$$

$$= \frac{0.67 \times 0.05}{0.67 \times 0.05 + 0.01 \times 0.95} = 0.78$$

Summarizing a PDF

- A PDF is a large table of numbers
- But generally, we don't need to know the entire thing; often the "highlights" are enough
  - *Expectation* and *Variance*
  - (statistics)

Expectation

- The expectation of r.v. X is defined as:

$$E(X) = \sum_{x} x p_X(x)$$

- The "average value" of X under $p_X(x)$

Expectation example

- A coin has 0.99 probability of showing heads. You get \$0 if the coin shows heads, and \$10 else. How much do you *expect to get* if I toss the coin?

$$E(X) = \sum_x x p_X(x) = (0 * 0.99 + 10 * 0.01) = \$0.1$$

Variance

- The variance of r.v. X is defined as:

$$V(X) = E([X - E(X)]^2)$$

$$= \sum_x (x - E(X))^2 p_X(x)$$

- The "average spread" of values *around the average* of the r.v.

Variance example

- A coin has 0.99 probability of showing heads. You get $0 if the coin shows heads, and $10 else. What is the variance of your takings?

$$E(X) = \sum_x x p_X(x) = (0*0.99 + 10*0.01) = \$0.1$$

$$V(X) = E([X - E(X)]^2)$$

$$= (0-0.1)^2 * 0.99 + (10-0.1)^2 * 0.01$$

$$= 0.99$$

Variance example

- A coin has 0.99 probability of showing heads. You get $10 if the coin shows heads, and $0 else. What is the variance of your takings?

$$E(X) = \sum_x x p_X(x) = (10*0.99 + 0*0.01) = \$9.9$$

$$\begin{aligned} V(X) &= E([X - E(X)]^2) \\ &= (10-9.9)^2 * 0.99 + (0-9.9)^2 * 0.01 \\ &= 0.99 \end{aligned}$$

Variance example 3

- A coin has 0.5 probability of showing heads. You get $0 if the coin shows heads, and $10 else. What is the variance of your takings?

$$E(X) = \sum_x x p_X(x) = (0*0.5 + 10*0.5) = \$5$$

$$\begin{aligned} V(X) &= E([X - E(X)]^2) \\ &= (0-5)^2 * 0.5 + (10-5)^2 * 0.5 \\ &= 25 \end{aligned}$$

Statistical Independence

- Two r.v.'s X and Y are statistically independent if

$$p_{X,Y}(X = x, Y = y) = p_X(X = x)\, p_Y(Y = y)$$

- If so, we can reason separately about x and y and then combine results---key factor in gaining efficiency (later)

Consequence

$$p_{X|Y}(X=x\,|\,Y=y)=\frac{p_{X,Y}(X=x,Y=y)}{p_Y(Y=y)}$$

$$=\frac{p_X(X=x)\,p_Y(Y=y)}{p_Y(Y=y)}$$

$$=p_X(X=x)$$

Conditional Independence

- Two r.v.'s X and Y are conditionally independent given a third, R, if

$$p_{X,Y|R}(X = x, Y = y \mid R = r) =$$
$$p_{X|R}(X = x \mid R = r)\, p_{Y|R}(Y = y \mid R = r)$$

Probabilistic Inference

- We are given a joint density function over a collection of random variables
- Based on this pdf, we want to find the probability of *some event, given observed variables*
  - Observed variables are evidence
  - Events are described by query variables

Example

CloudyTomorrow	RainTomorrow	WetGrass	Probability
No	No	No	0.4
No	No	Yes	0.01
No	Yes	No	0
No	Yes	Yes	0.01
Yes	No	No	0.15
Yes	No	Yes	0.02
Yes	Yes	No	0.01
Yes	Yes	Yes	0.4

$$\Pr(WetGrass = Yes \mid CloudyTomorrow = Yes)?$$

Query variable

evidence

Inference by Enumeration

- We are given a pdf over a collection of r.v.'s $\mathbf{X}$
  - Of these, we observe evidence $\mathbf{E}=\mathbf{e}$ ($\mathbf{E} \subseteq \mathbf{X}$)
  - We are interested in the query variable V
  - Let $\mathbf{Y}$ be $\mathbf{X} \setminus \{\mathbf{E}, V\}$ (everything in $\mathbf{X}$ not in $\mathbf{E}$ and not V)
    - Note $\mathbf{X} = \mathbf{Y} \cup \mathbf{E} \cup V$
    - Sometimes called "nuisance" variables

- We want $p(V=v|\mathbf{E}=\mathbf{e})$

Inference by Enumeration

$$p(V=v\,|\,\mathbf{E}=\mathbf{e}) = \frac{p(V=v,\mathbf{E}=\mathbf{e})}{p(\mathbf{E}=\mathbf{e})}$$

Marginalization

$$p(V=v,\mathbf{E}=\mathbf{e}) = \sum_{\mathbf{y}} p(V=v,\mathbf{E}=\mathbf{e},\mathbf{Y}=\mathbf{y}),\ \mathbf{Y}=\mathbf{X}\setminus\{\mathbf{E},V\}$$

Atomic Event

$$p(\mathbf{E}=\mathbf{e}) = \sum_{v}\sum_{\mathbf{y}} p(V=v,\mathbf{E}=\mathbf{e},\mathbf{Y}=\mathbf{y})$$

Normalization Factor

$$p(V=v\,|\,\mathbf{E}=\mathbf{e}) = \frac{\sum_{\mathbf{y}} p(V=v,\mathbf{E}=\mathbf{e},\mathbf{Y}=\mathbf{y})}{\sum_{v}\sum_{\mathbf{y}} p(V=v,\mathbf{E}=\mathbf{e},\mathbf{Y}=\mathbf{y})}$$