EECS 391: Introduction to AI

Soumya Ray

Website: http://vorlon.case.edu/~sray/eecs391_sp12/index.html

Email: sray@case.edu

Office: Olin 516

Office hours: Tue 2:30-4:00 or by appointment

Announcements

- HW4 due, solutions tomorrow noon
- Quiz 4 Thursday
- I will have office hours next week at same time

Today

- Part 3: Probability and Machine Learning
  - Basic Probability and Statistics (Ch 13)
  - Note: this is meant to be a refresher rather than a mathematically rigorous introduction
  - Please review your old probability textbooks if too many cobwebs

Basic Concepts

Representing Uncertainty

- We have seen logical representations and how they can be used to solve problems
- But the real world is uncertain due to partial observability, cost of acquiring information, limits on computation etc
- How to represent/reason in this case?

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

Example

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

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

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
  - (Similar to a row in the 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 built out of 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) ("truth table")

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
  - $\Pr(RainTomorrow=Yes, CloudyTomorrow=Yes) = 0.99$
  - $\Pr(-1 \leq x \leq 1, -1 \leq y \leq 1) = 0.2$

- This is called a "joint probability"

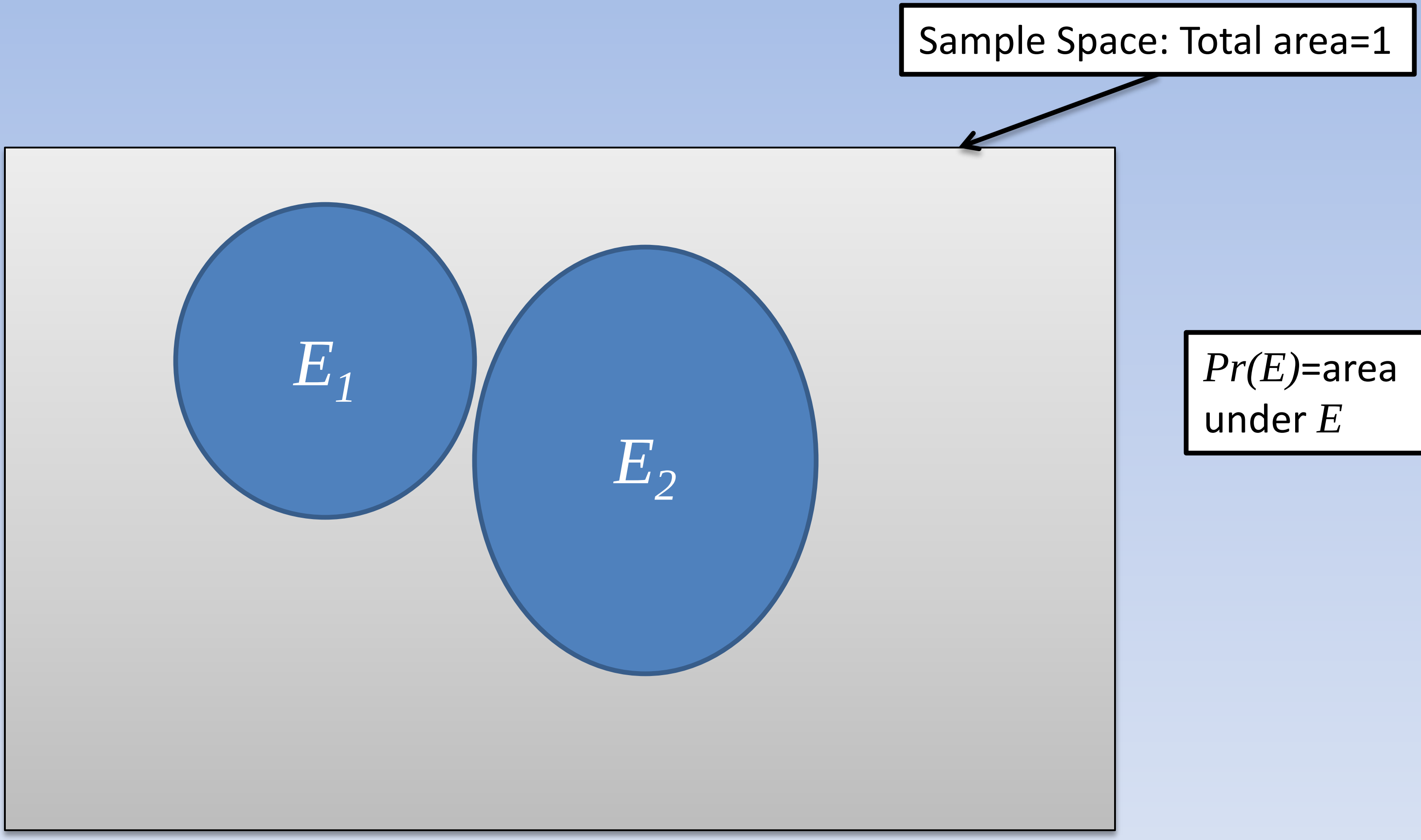
Axioms of Probability

- For any event $E, 0 \leq \Pr(E) \leq 1$
- The probability of the sample space is 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.

Pictorial Representation



$Pr(E)$=area under 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). (Bruno de Finetti 1931)

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 $Pr(\bar{E})$?

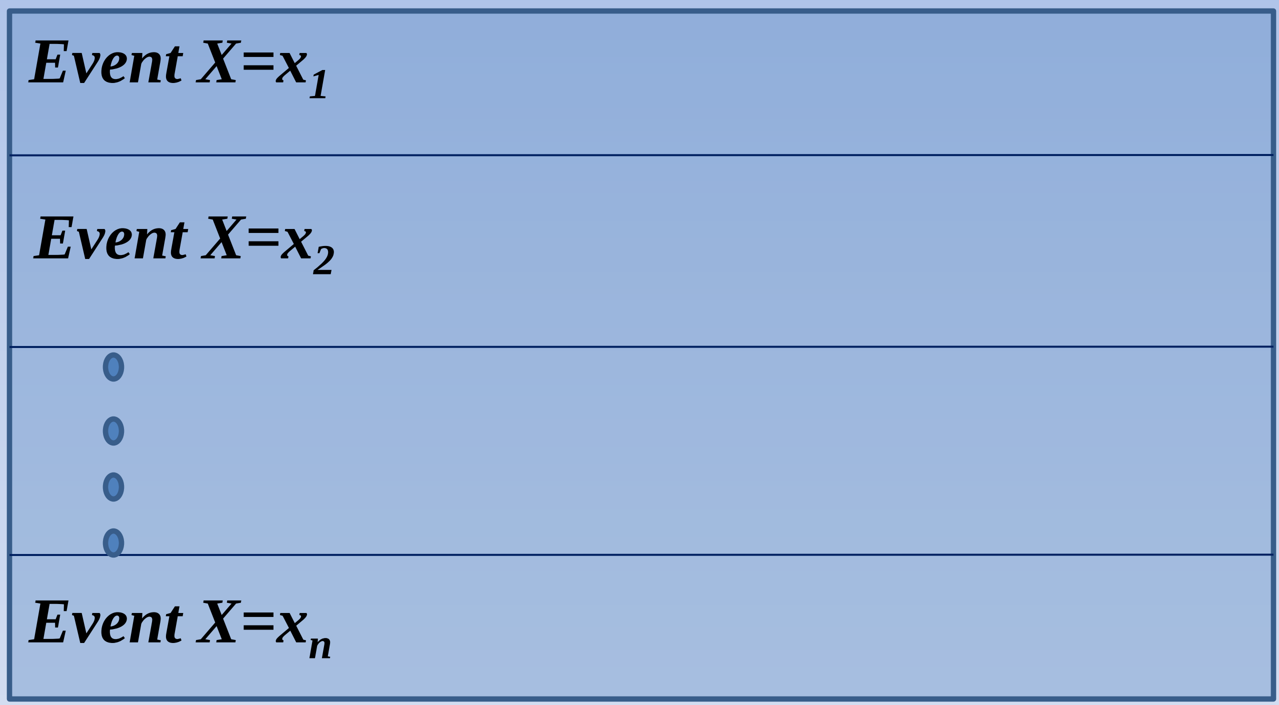
Probability Density Functions

- 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.)

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

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

PDFs must sum to 1



Joint PDF

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

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

First Order Logic

- Is it possible to augment FOL in the same way?
  - Hot topic of current research!

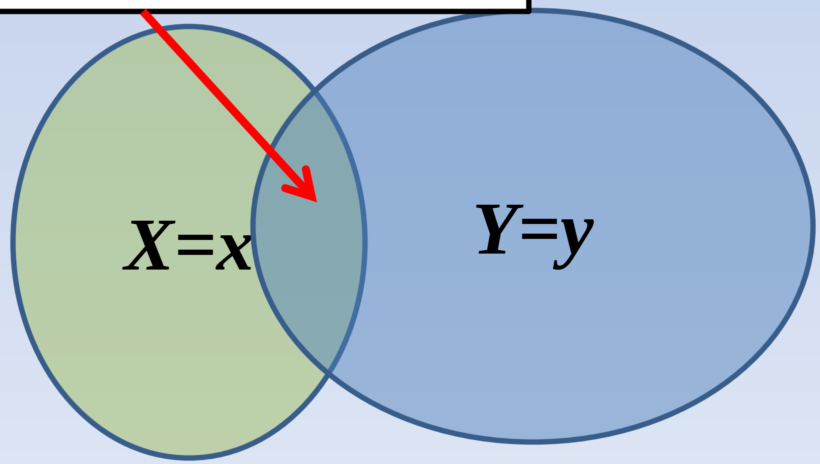
Terminology and Results

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

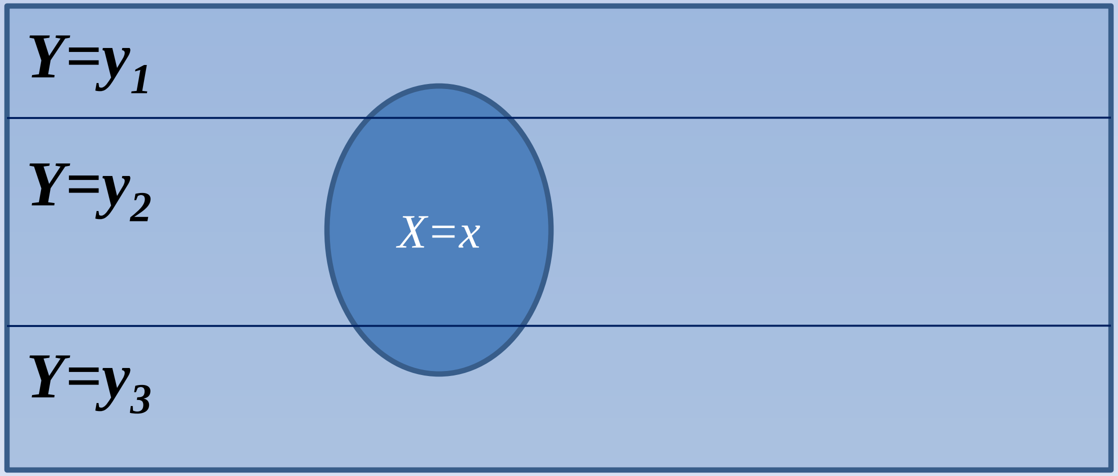
- From the definition of conditional probability:

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

Marginalization

- For any two random variables X and Y:

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



Conditioning

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

Marginalization

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

Product Rule

Bayes' Rule

(Rev. Thomas Bayes 1763)

- Suppose E and C are two r.v.'s.
- Then

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

Very important! Remember this.

Proof

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

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**---fundamental to *learning*

Example

- Lung cancer can be caused by smoking or by a genetic defect. 15% of the population are smokers. 1 in 3 who smoke and 1 in 100 who don’t get the disease.
- Suppose X has the disease. What is the probability X does not smoke?

Statistical Independence

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

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

Consequence

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

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

$$= p_X(X = x)$$

Conditional Independence

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

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

Expectation

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

$$E(X) = \sum_x x p_X(x)$$

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

- A coin has 0.9 probability of showing heads. Let r.v. X take value 0 if the coin shows heads, and 1 else. What is $E(X)$?

Expectation contd

- For any real-valued function *g(X)*:

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

Expectation Properties

- For any two r.v.'s X and Y,

  $$E(X+Y)=E(X)+E(Y)$$

  – Expectation is a "linear operator"

- For any two independent r.v.'s X and Y,

  $$E(XY)=E(X)E(Y)$$

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 of a r.v. around the average of the r.v.

Variance Properties

- For any two independent r.v.'s X and Y:
  $V(X+Y)=V(X)+V(Y)$

Tasks

- There are two general tasks we will be interested in when working with random variables and their associated distributions
  - Inference
  - Estimation