CS7480: Topics in Programming Languages: Probabilistic Programming

Lecture 3: Probability & Logic Continued

Instructor: Steven Holtzen

Place: Northeastern University

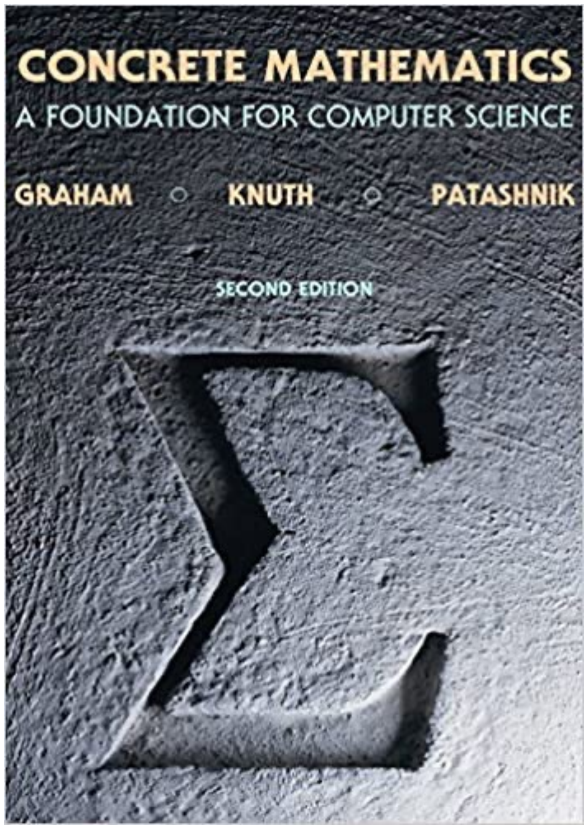
Term: Fall 2021

Course webpage:
https://www.khoury.northeastern.edu/home/sholtzen/CS7480Fall21/

Overview

- Last time we learned…
  - Syntax, grammars, and parse trees
  - Formal semantics for propositional logic $[\![p]\!]v$
  - Search trees, SAT
  - Why? These will be important foundations for *program reasoning*
- Today…
  - Pick up from the beginning of probability (I will give simpler definitions this time)
  - Work towards probabilistic logic, our first “probabilistic programming language”

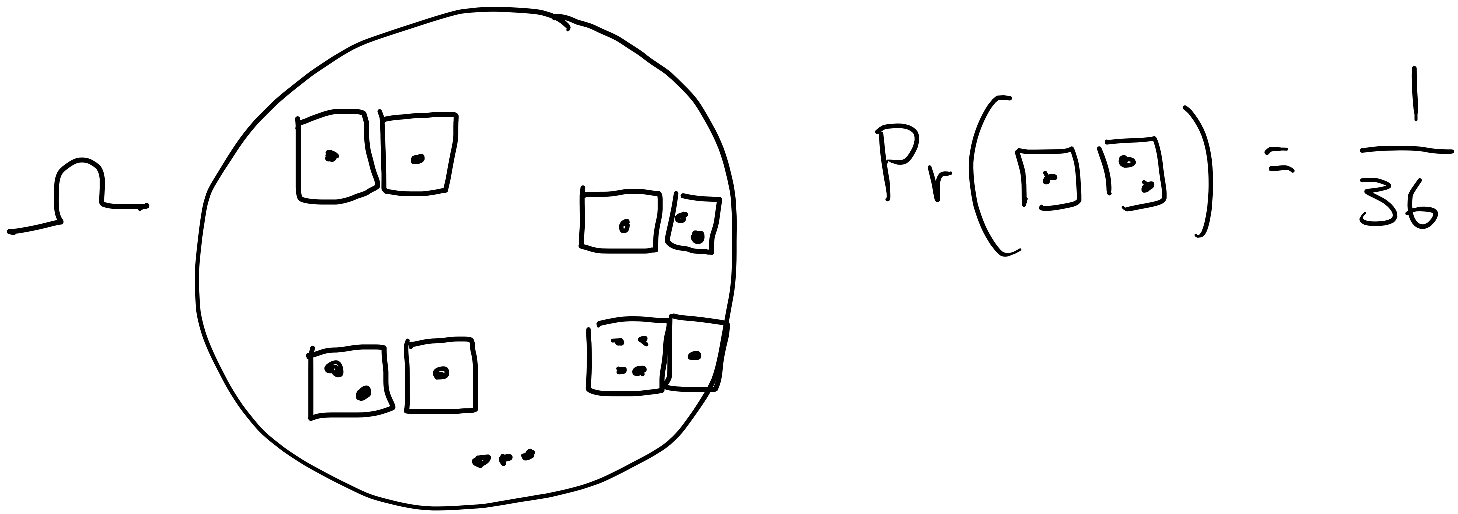
I’m following a book



- Chapter 8, “Discrete Probability”
- Many extremely good exercises at the end of the chapter (with solutions in the back of the book); give them a try to brush up!
- If you’re feeling shaky, after lecture try exercise 2, 8, 9

Probability Space

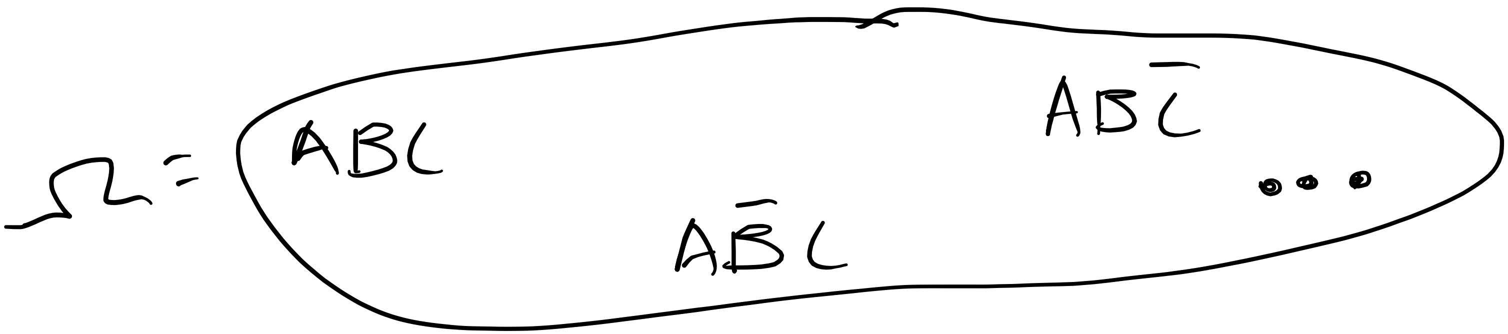
- Written Ω, it is the *set of all things that can happen*



- Paired with a *probability distribution* $Pr: \Omega \to [0,1]$
  - Must satisfy *unit measure* $\sum_{\omega \in \Omega} \Pr(\omega) = 1$.

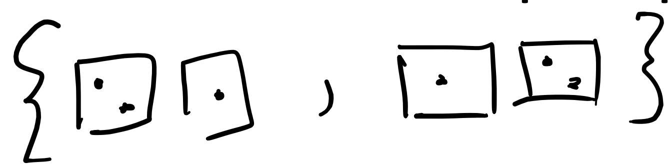
Probability Space on Assignments

- Let $V = \{A, B, C\}$ be a set of propositional variables
- Then we can define a sample space on assignments to V:

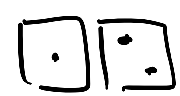
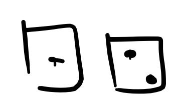


- We can define a probability distribution on this however we want

Events

- An *event* E is a subset of the sample space, $E \subseteq \Omega$
  - "Sums to 3": 
- The *probability of an event* is defined

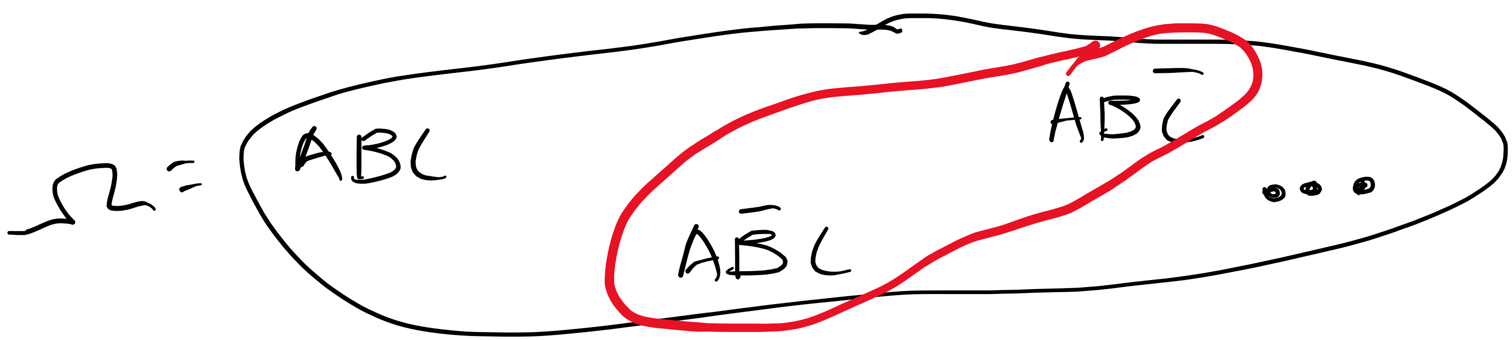
$$\Pr(E) = \sum_{\omega \in E} \Pr(\omega).$$

Pr(Sums to 3) = Pr() + Pr()

Finite additivity holds by definition

Probability of Satisfaction

- Suppose we have a logical sentence φ; this naturally defines an *event* on the sample space of assignments $E_\varphi = \{\, \omega \in \Omega \mid \omega \vDash \varphi \,\}$



- Then, $\Pr(E_\varphi)$ is the probability that φ is satisfied by a world drawn according to some distribution on Ω

Random Variables

- A *random variable* is a function on the sample space $X: \Omega \to \mathbb{R}$ (usually denoted w/ capital letters)
    - Example: let S be a function that sums dice rolls

$S($⚀⚃$) = 5$

- Probability of a random variables:

$\Pr(S=4) = \Pr(\{$⚁⚁, ⚀⚂, ⚂⚀$\})$

Random Variables: Derived Distributions

- We could have defined the "sums" distribution directly using a sample space $\Omega = \{2, 3, \ldots, 12\}$

$\Pr(2)$	$\Pr(3)$	$\Pr(4)$	…
1/36	2/36	3/36	

Tabular representation

- The "sums" distribution is *derived* from the "dice pairs" distribution

Expectations

- The “average value” of a random variable

$$\mathbb{E}[X] = \sum_{(x \in X)} \Pr(x) \times x$$

- Expectations have many nice algebraic properties

Joint Distributions

- Why are random variables interesting? Typically, if we have *more than one derived from the same sample space*

$$\Pr(X = x, Y = y)$$

Called *joint probability distribution over* X and Y

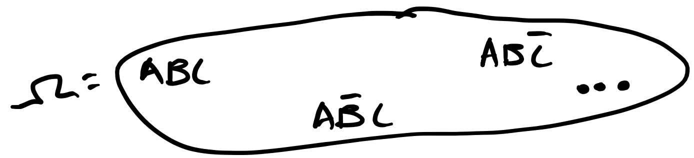
- Ex: On $\Omega =$set of dice rolls
  - $S =$sum, $S_1 =$ first dice roll, $S_2 =$second dice roll
- Then $\Pr(S_1, S_2) = \Pr(S_1) \times \Pr(S_2)$

Very common shorthand for $\Pr(S_1 = s_1)$

  - Called *independent random variables*
- On the other hand, $\Pr(S, S_1) \neq \Pr(S) \times \Pr(S_1)$

Joint Distribution on Dice Rolls

Joint Distributions on Propositional Worlds

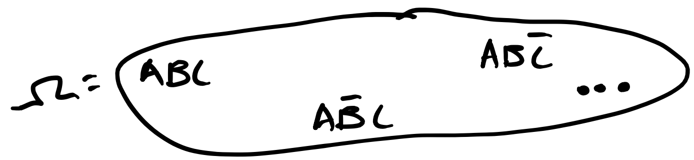


- Let $E_A = \{\omega \in \Omega \mid A = \mathrm{T} \text{ in } \omega\}$ for any variable A
- Then, $\Pr(E_A, E_B) = \Pr(E_A) \times \Pr(E_B)$ *may not hold* for arbitrary distributions on Ω
    - Exercise: find an example!

Queries & Inference

Queries?

- So we have defined this great distribution that describes the world

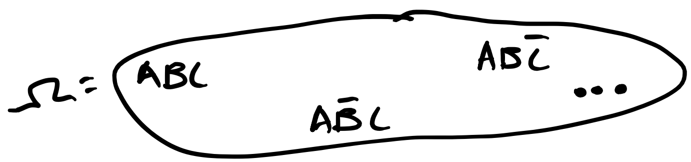


$+ \Pr(\omega)$

- What do we want to do with it?
  - Ask it questions! This is *querying*

Query: Probability of Evidence

- Given a distribution…



$+ \Pr(\omega)$

- Compute the probability of some event E
- Wait! Isn't this easy? It's just the definition…

$$\Pr(E) = \sum_{\omega \in E} \Pr(\omega)$$

Complexity of Querying

- So far we've been working under a *tabular representation* of a distribution

  $\Omega = \{2, 3, \dots, 12\}$

$\Pr(2)$	$\Pr(3)$	$\Pr(4)$	$\Pr(5)$	$\Pr(6)$	$\Pr(7)$	$\Pr(8)$	$\Pr(9)$	$\Pr(10)$	$\Pr(11)$	$\Pr(12)$
1/36	2/36	3/36	4/36	5/36	6/36	5/36	4/36	3/36	2/36	1/36

- With this representation, computing the probability of evidence (for arbitrary queries) requires *scanning the table and summing each entry*
  - $O(|\Omega|)$ evaluates of $\Pr(\omega)$; this is usually our scaling parameter of interest

Complexity of Querying

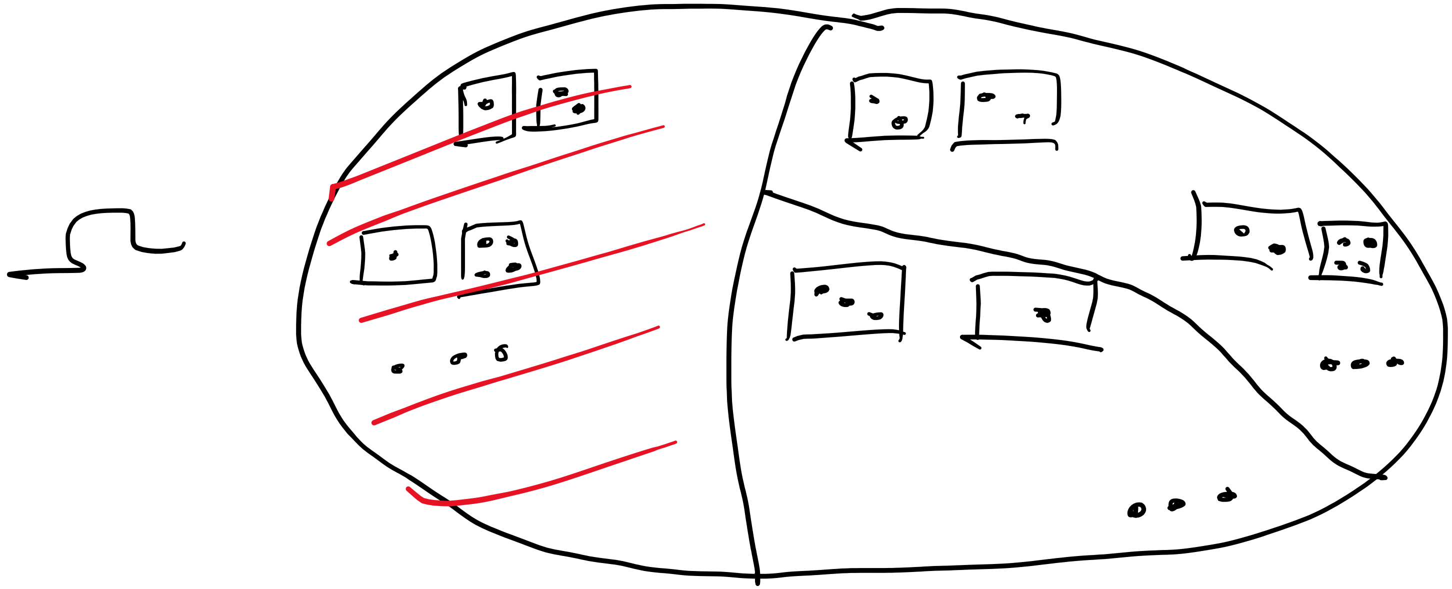
- The complexity of various queries is intimately related to *how you represent the probability distribution*

- For now, we will continue considering tables

Query: Conditioning

- "Given that I see an even dice roll, what is the probability that one of the dice rolls was a 2?"

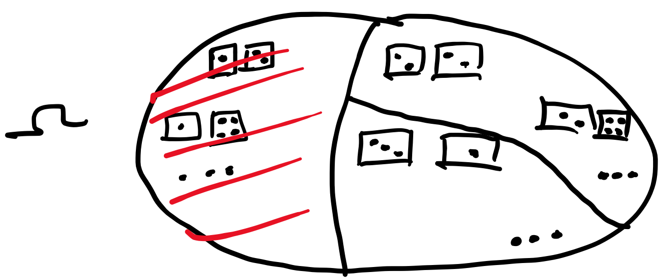
$$\Pr(D_1 = 2 \mid S \text{ is even})$$



Conditioning

- "Given that I see an even dice roll, what is the probability that one of the dice rolls was a 2?"

$$\Pr(D_1 = 2 \mid S \text{ is even})$$



$$\Pr(X \mid E) = \frac{\Pr(X \wedge E)}{\Pr(E)}$$

Normalizing Constant / Probability of Evidence

Turned into 2 evidence probability queries

Conditioning

- $\Pr(\text{first is even} \mid \text{sum is even})$?

$\Pr(2)$	$\Pr(3)$	$\Pr(4)$	$\Pr(5)$	$\Pr(6)$	$\Pr(7)$	$\Pr(8)$	$\Pr(9)$	$\Pr(10)$	$\Pr(11)$	$\Pr(12)$
1/36	2/36	3/36	4/36	5/36	6/36	5/36	4/36	3/36	2/36	1/36

Conditioning and Independence

- X and Y are independent if and only if $\Pr(X \mid Y) = \Pr(X)$

$$\Pr(X \mid Y) = \frac{\Pr(X \wedge Y)}{\Pr(Y)} = \frac{\Pr(Y) \times \Pr(X)}{\Pr(Y)} = \Pr(X) .$$

Query: Marginalization

- Given a joint distribution $\Pr(X, Y)$, the *marginal distribution on* X is:

$$\Pr(X) = \sum_{y} \Pr(X, Y) \, .$$

Query: Optimization

- What is the *most likely state (given evidence E)?*
$$argmax_{\omega} \Pr(\omega \mid E)?$$
  - Called *Most Probable Explanation (MPE)*
  - Again, requires scanning the whole table

$\Pr(2)$	$\Pr(3)$	$\Pr(4)$	$\Pr(5)$	$\Pr(6)$	$\Pr(7)$	$\Pr(8)$	$\Pr(9)$	$\Pr(10)$	$\Pr(11)$	$\Pr(12)$
1/36	2/36	3/36	4/36	5/36	6/36	5/36	4/36	3/36	2/36	1/36

Conclusion

- Logic gives a syntax for describing a set of worlds
  - There is syntax and semantics
  - There are analyses (satisfiability)
  - Search and decision trees

- Probability describes distributions on those worlds