Discrete Structures for Computer Science

William Garrison
bill@cs.pitt.edu
6311 Sennott Square

Lecture #23: Discrete Probability

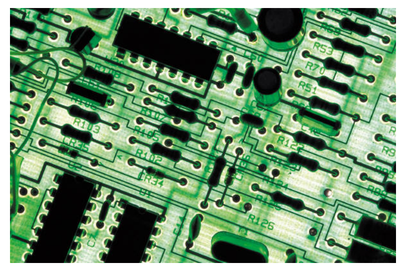
University of Pittsburgh

Based on materials developed by Dr. Adam Lee

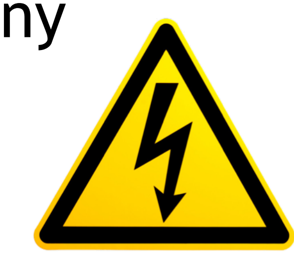
The study of probability is concerned with the likelihood of events occurring

Like combinatorics, the origins of probability theory can be traced back to the study of gambling games

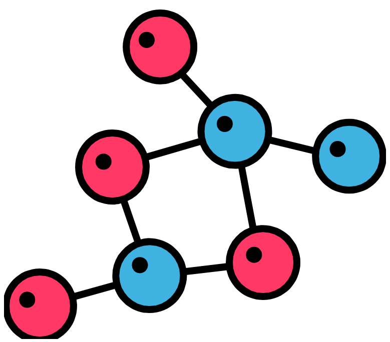
Still a popular branch of mathematics with many applications:



Simulation



Risk Assessment



Genetics

Algorithm Design

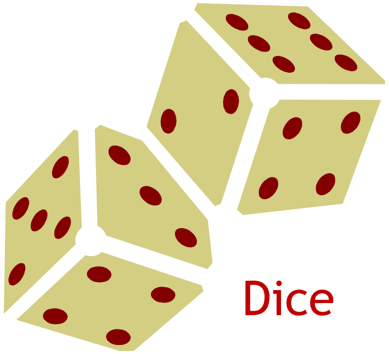
Gambling

Many situations can be analyzed using a simplified model of probability

Assumptions:

1. Finite number of possible outcomes
2. Each outcome is equally likely

Card games


Dice


Flipping coins


Roulette

Lotteries

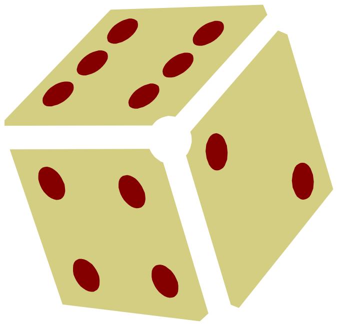
Terminology

Definitions:

- An experiment is a procedure that yields one of a given set of possible outcomes
- The sample space of an experiment is the set of possible outcomes
- An event is a subset of the sample space
- Given a finite sample space S of equally-likely outcomes, the probability of an event E is p(E) = |E|/|S|.

Example:

- Experiment: Roll a single 6-sided die one time
- Sample space: {1, 2, 3, 4, 5, 6}
- One possible event: Roll an even number ⇒ {2, 4, 6}
- The probability of rolling an even number is |{2, 4, 6}| / |{1, 2, 3, 4, 5, 6}| = 3/6 = 1/2



Solving these simplified finite probability problems is "easy"

I told you that combinatorics and probability were related!

Step 1: Identify and count the sample space

Step 2: Count the size of the desired event space

Step 3: Divide!

When two dice are rolled, what is the probability that the sum of the two numbers is seven?

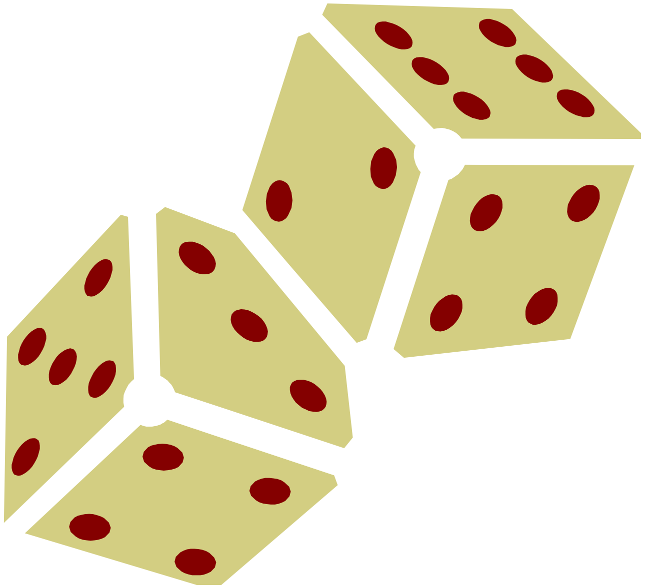
Step 1: Identify and count sample space

- Sample space, S, is all possible pairs of numbers 1-6
- Product rule tells us that $|S| = 6^2 = 36$

Step 2: Count event space

- (1, 6)
- (2, 5)
- (3, 4)
- (4, 3)
- (5, 2)
- (6, 1)

$|E| = 6$



Step 3: Divide

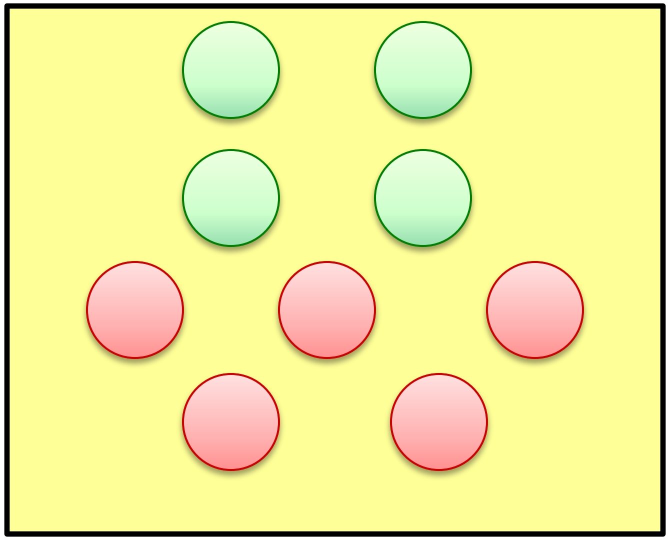
- Probability of rolling two dice that sum to 7 is p(E)
- $p(E) = |E| / |S| = 6/36 = 1/6$

Balls and Bins

Example: A bin contains 4 green balls and 5 red balls. What is the probability that a ball chosen from the bin is green?

Solution:

- 9 possible outcomes (balls)
- 4 green balls, so $|E| = 4$
- So $p(E) = 4/9$ that a green ball is chosen



Hit the lotto

Example: Suppose a lottery gives a large prize to a person who picks 4 digits between 0-9 in the correct order, and a smaller prize if only three digits are matched. What is the probability of winning the large prize? The small prize?

Solution:

Grand prize

- S = possible lottery outcomes
- $|S| = 10^4 = 10,000$
- E = all 4 digits correct
- $|E| = 1$
- So $p(E) = 1/10,000 = 0.0001$

Smaller prize

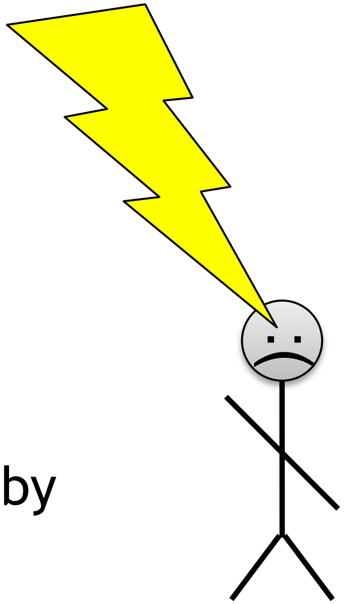
- S = possible lottery outcomes
- $|S| = 10^4 = 10,000$
- E = one digit incorrect
- We can count $|E|$ using the sum rule:
  - 9 ways to get 1st digit wrong OR
  - 9 ways to get 2nd digit wrong OR
  - 9 ways to get 3rd digit wrong OR
  - 9 ways to get 4th digit wrong
- So $|E| = 9 + 9 + 9 + 9 = 36$
- $p(E) = 36/10,000 = 0.0036$

Mega Lotteries

Example: Consider a lottery that awards a prize if a person can correctly choose a set of 6 numbers from the set of the first 40 positive numbers. What is the probability of winning this lottery?

Solution:

- S = All sets of six numbers between 1 and 40
- Note that order does not matter in this lottery
- Thus, $|S| = C(40, 6) = 40!/(6!34!) = 3{,}838{,}380$
- Only one way to do this correctly, so $|E| = 1$
- So $p(E) = 1/3{,}838{,}380 \approx 0.00000026$



Lesson: You stand a better chance at being struck by lightning than winning this lottery!

Four of a Kind

Example: What is the probability of getting "four of a kind" in a 5-card poker hand?

Solution:

- S = set of all possible poker hands
- Recall |S| = C(52,5) = 2,598,960
- E = all poker hands with 4 cards of the same type
- To draw a four of a kind hand:
  - C(13, 1) ways to choose the type of card (2, 3, …, King, Ace)
  - C(4,4) = 1 way to choose all 4 cards of that type
  - C(48, 1) ways to choose the 5^{th} card in the hand
  - So, |E| = C(13,1)C(4,4)C(48,1) = 13 $\times$ 48 = 624
- p(E) = 624/2,598,960 ≈ 0.00024

A Full House

Example: What is the probability of drawing a full house when drawing a 5-card poker hand? (Reminder: A full house is three cards of one kind, and two cards of another kind.)

Solution:

- $|S| = C(52,5) = 2,598,960$
- E = all hands containing a full house
- To draw a full house:
  - Choose two types of cards (order matters)
  - Choose three cards of the first type
  - Choose two cards of the second type
- So $|E| =$
- $p(E) =$

Sampling with or without replacement makes a difference!

Example: Consider a bin containing balls labeled with the numbers 1, 2, …, 50. How likely is the sequence 23, 4, 3, 12, 48 to be drawn in order if a selected ball is not returned to the bin? What if selected balls are immediately returned to the bin?

Solution:

- Note: Since order is important, we need to consider 5-permutations
- If balls are not returned to the bin, we have $P(50, 5) = 50 \times 49 \times 48 \times 47 \times 46 = 254{,}251{,}200$ ways to select 5 balls
- If balls are returned, we have $50^5 = 312{,}500{,}000$ ways to select 5 balls
- Since there is only one way to select the sequence 23, 4, 3, 12, 48 in order, we have that
  - $p(E) = 1/254{,}251{,}200$ if balls are not replaced
  - $p(E) = 1/312{,}500{,}000$ if balls are replaced

Yes, calculating probabilities can be easy

Anyone can divide two numbers!

But, Be careful when you:

- Define the sets S and E
- Count the cardinality of S and E

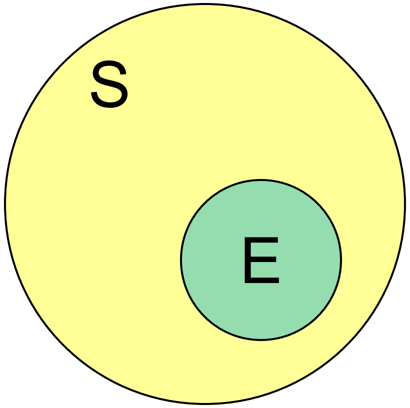
In-class exercises

Problem 1: Consider a box with 3 green balls and 1 pink ball. What is the probability of drawing a pink ball? What is the probability of drawing two green balls in two successive picks (without replacement)?

Problem 2: In poker, a straight flush is a hand in which all 5 cards are from the same suit and occur in order. For example, a hand containing the 3, 4, 5, 6, and 7 of hearts would be a straight flush, while the hand containing the 3, 4, 5, 7, and 8 of hearts would not be. Note that a royal flush (10 through A) is not considered a straight flush (but A through 5 is). What is the probability of drawing a straight flush in poker?

Problem 3: A flush is a hand in which all five cards are of the same suit, but do not form an ordered sequence. What is the probability of drawing a flush in poker?

What about events that are derived from other events?

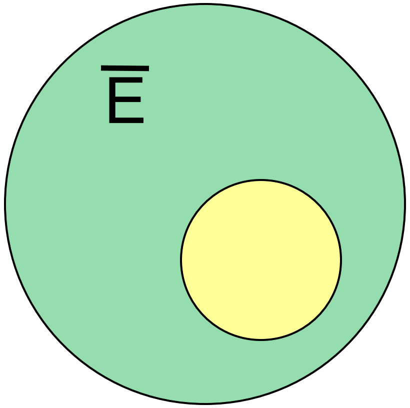


Recall: An event E is a subset of the sample space S

Definition: $p(\overline{E}) = 1 - p(E)$

Proof:

- Note that $\overline{E} = S - E$, since S is universe of all possible outcomes
- So, $|\overline{E}| = |S| - |E|$
- Thus, $p(\overline{E}) = |\overline{E}|/|S|$ by definition
- $= (|S| - |E|)/|S|$ by substitution
- $= 1 - |E|/|S|$ simplification
- $= 1 - p(E)$ ❑ by definition



Why is this useful?

Sometimes, counting |E| is hard!

Example: A 10-bit sequence is randomly generated. What is the probability that at least 1 bit is 0?

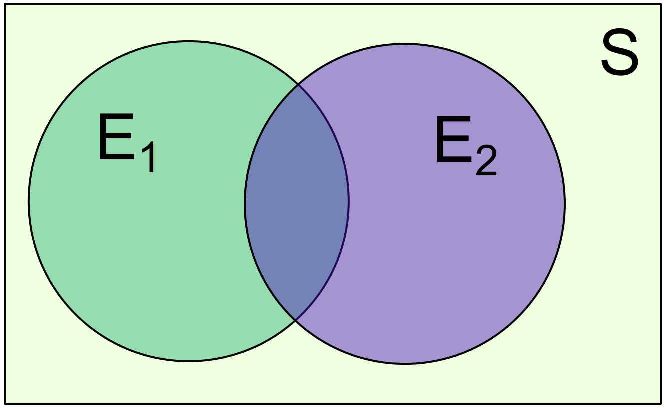
Solution:

- S = all 10-bit strings
- $|S| = 2^{10}$
- E = all 10-bit strings with at least 1 zero
- $\overline{E}$ = all 10-bit strings with no zeros = {1111111111}
- $p(E) = 1 - p(\overline{E})$
- $= 1 - 1/2^{10}$
- $= 1 - 1/1024$
- $= 1023/1024$

So the probability of a randomly generated 10-bit string containing at least one 0 is 1023/1024.

We can also calculate the probability of the union of two events

Definition: If E_1 and E_2 are two events in the sample space S, then $p(E_1 \cup E_2) = p(E_1) + p(E_2) - p(E_1 \cap E_2)$.



Why does this look familiar?

Proof:

- Recall: $|E_1 \cup E_2| = |E_1| + |E_2| - |E_1 \cap E_2|$
- $p(E_1 \cup E_2) = |E_1 \cup E_2| / |S|$
- $= (|E_1| + |E_2| - |E_1 \cap E_2|) / |S|$
- $= |E_1|/|S| + |E_2|/|S| - |E_1 \cap E_2|/|S|$
- $= p(E_1) + p(E_2) - p(E_1 \cap E_2)$

Divisibility…

Example: What is the probability that a positive integer not exceeding 100 is divisible by either 2 or 5?

Solution:

- Let E_1 be the event that an integer is divisible by 2
- Let E_2 be the event that an integer is divisible by 5
- $E_1 \cup E_2$ is the event that an integer is divisible by 2 or 5
- $E_1 \cap E_2$ is the event that an integer is divisible by 2 and 5
- $|E_1| = 50$
- $|E_2| = 20$
- $|E_1 \cap E_2| = 10$
- $p(E_1 \cup E_2) =$
- $=$
- $=$
- $=$

Not all events are equally likely to occur…

Sporting events

Investments

Games of strategy

Nature

We can model these types of real-life situations by relaxing our model of probability

As before, let S be our sample space. Unlike before, we will allow S to be either finite or countable.

We will require that the following conditions hold:

1. $0 \le p(s) \le 1$ for each $s \in S$
2. $\sum_{s \in S} p(s) = 1$

No event can have a negative likelihood of occurrence, or more than a 100% chance of occurrence

In 100% of experiments, one of the events occurs

The function $p : S \rightarrow [0,1]$ is called a probability distribution

Simple example: Fair and unfair coins

Example: What probabilities should be assigned to outcomes heads (H) and tails (T) if a fair coin is flipped? What if the coin is biased so that heads is twice as likely to occur as tails?

Case 1: Fair coins

- Each outcome is equally likely
- So p(H) = 1/2, p(T) = 1/2
- Check:
  - $0 \le 1/2 \le 1$
  - $1/2 + 1/2 = 1$

✔

Case 2: Biased coins

- Note:
  1. p(H) = 2p(T)
  2. p(H) + p(T) = 1
- 2p(T) + p(T) = 1
- 3p(T) = 1
- p(T) = 1/3, p(H) = 2/3

Are the following probability distributions valid? Why or why not?

$S = \{1, 2, 3, 4\}$ where

- $p(1) = 1/3$
- $p(2) = 1/6$
- $p(3) = 1/6$
- $p(4) = 1/3$

$S = \{1, 2, 3, 4\}$ where

- $p(1) = 2/3$
- $p(2) = 1/6$
- $p(3) = -1/6$
- $p(4) = 1/3$

$S = \{a, b, c\}$

- $p(a) = 3/4$
- $p(b) = 1/4$
- $p(c) = 0$

$S = \{a, b, c\}$

- $p(a) = 1/2$
- $p(b) = 1/4$
- $p(c) = 0$

More definitions

Definition: Suppose that S is a set with n elements. The uniform distribution assigns the probability 1/n to each element of S.

The distribution of fair coin flips is a uniform distribution!

Definition: The probability of an event $E \subseteq S$ is the sum of the probabilities of the outcomes in E. That is:

$$p(E) = \sum_{s \in E} p(s)$$

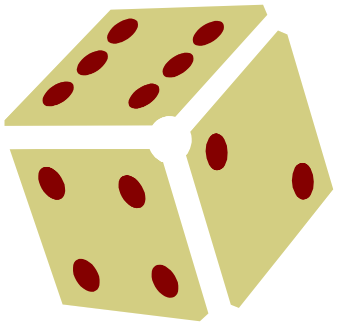
Loaded dice

Example: Suppose that a die is biased so that 3 appears twice as often as each other number, but that the other five outcomes are equally likely. What is the probability that an odd number appears when we roll this die?

Solution:

- $p(1) + p(2) + p(3) + p(4) + p(5) + p(6) = 1$
- Note that $p(1) = p(2) = p(4) = p(5) = p(6)$ and $p(3) = 2p(1)$
- So, $p(1) + p(1) + 2p(1) + p(1) + p(1) + p(1) = 7p(1) = 1$
- Thus $p(1) = p(2) = p(4) = p(5) = p(6) = 1/7$ and $p(3) = 2/7$

- Now, we want to find $p(E)$, where $E = \{1, 3, 5\}$
- $p(E) = p(1) + p(3) + p(5)$
- $\quad = 1/7 + 2/7 + 1/7$
- $\quad = 4/7$



In-class exercises

Consider a die in which (i) 1, 3, and 4 are rolled with the same frequency, (ii) 2 is rolled 3 times as often as 1, (iii) 5 is rolled 2 times as often as 4, (iv) and 6 is rolled 4 times as often as 2.

Problem 4: What is the probability distribution for this die?

Problem 5: What is the probability of rolling a 1 or a 3?

Problem 6: What is the probability of rolling an even number? An odd number?

Final Thoughts

- Probability allows us to analyze the likelihood of events occurring

- Today, we learned how to analyze events that are equally likely, as well as those that have non-equal probabilities of occurrence

- Next time:
  - More probability theory (Section 6.2)