Discrete Structures for Computer Science

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

Lecture #21: Discrete Probability

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:

- Genetics
- Algorithm Design
- Simulation
- Risk Assessment
- 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
**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 $p(E) = \frac{|E|}{|S|}$.

**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 $\Rightarrow \{2, 4, 6\}$
- The probability of rolling an even number is $\frac{|\{2, 4, 6\}|}{|\{1, 2, 3, 4, 5, 6\}|} = \frac{3}{6} = \frac{1}{2}$
Solving these simplified finite probability problems is “easy”

Step 1: Identify and count the sample space

Step 2: Count the size of the desired event space

Step 3: Divide!

I told you that combinatorics and probability were related!
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 $p(E)$
- $p(E) = \frac{|E|}{|S|} = \frac{6}{36} = \frac{1}{6}$
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
- \( S = \text{possible lottery outcomes} \)
- \( |S| = 10^4 = 10,000 \)
- \( E = \text{all 4 digits correct} \)
- \( |E| = 1 \)
- So \( p(E) = 1/10,000 = 0.0001 \)

Smaller prize
- \( S = \text{possible lottery outcomes} \)
- \( |S| = 10^4 = 10,000 \)
- \( E = \text{one digit incorrect} \)
- 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 \( |E| = 9 + 9 + 9 + 9 = 36 \)
- \( p(E) = 36/10,000 = 0.0036 \)
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 rank
- To draw a four of a kind hand:
  - $C(13, 1)$ ways to choose the rank of card (2, 3, ..., King, Ace)
  - $C(4,4) = 1$ way to choose all 4 cards of that rank
  - $C(48, 1)$ ways to choose the 5th card in the hand
  - So, $|E| = C(13,1)C(4,4)C(48,1) = 13 \times 48 = 624$
- $p(E) = 624/2,598,960 \approx 0.00024$
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 rank, and two cards of another rank.)

Solution:

- $|S| = \binom{52}{5} = 2,598,960$
- $E =$ all hands containing a full house
- To draw a full house:
  - Choose two ranks of cards (order matters)
  - Choose three cards of the first rank
  - Choose two cards of the second rank
- So $|E| =$
- $p(E) =$
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:** $p(E) = 1 - p(E)$

**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
- $p(\overline{E}) = (|S| - |E|) / |S|$ by substitution
- $p(\overline{E}) = 1 - |E| / |S|$ simplification
- $p(\overline{E}) = 1 - p(E)$ 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:**

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

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

$$p(E_1 \cup E_2) = p(E_1) + p(E_2) - p(E_1 \cap E_2).$$

**Proof:**

- **Recall:** $|E_1 \cup E_2| = |E_1| + |E_2| - |E_1 \cap E_2|$
- $p(E_1 \cup E_2) = \frac{|E_1 \cup E_2|}{|S|}$
- $= \frac{(|E_1| + |E_2| - |E_1 \cap E_2|)}{|S|}$
- $= \frac{|E_1|}{|S|} + \frac{|E_2|}{|S|} - \frac{|E_1 \cap E_2|}{|S|}$
- $= p(E_1) + p(E_2) - p(E_1 \cap E_2)$

Why does this look familiar?
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
- $|E_1| = 50$
- $|E_2| = 20$
- $|E_1 \cap E_2| = 10$
- $p(E_1 \cup E_2) = \frac{50}{100} + \frac{20}{100} - \frac{10}{100} = \frac{1}{2} + \frac{1}{5} - \frac{1}{10} = \frac{3}{5}$
A formula for the probability of pairwise disjoint events

**Theorem:** If $E_1$, $E_2$, ..., $E_n$ is a sequence of pairwise disjoint events in a sample space $S$, then we have:

$$p \left( \bigcup_{i=1}^{n} E_i \right) = \sum_{i=1}^{n} p(E_i)$$

Recall: $E_1$, $E_2$, ..., $E_n$ are pairwise disjoint iff $E_i \cap E_j = \emptyset$ for $1 \leq i,j \leq n$

We can prove this theorem using mathematical induction!
Final Thoughts

- Probability allows us to analyze the likelihood of events occurring

- Probability of some events can be derived from that for other events

- Today, we learned how to analyze events that are equally likely; next, we will analyze events that are not equally likely

- Next time:
  - Probability theory (Section 7.2)
The proof...

\[ P(n) \equiv p(\bigcup_{i=1}^{n} E_i) = \sum_{i=1}^{n} p(E_i) \]

**Base case:** \( P(2) \): Let \( E_1, E_2 \) be disjoint events.
- By definition, \( p(E_1 \cup E_2) = p(E_1) + p(E_2) - p(E_1 \cap E_2) \).
- Since \( E_1 \cap E_2 = \emptyset \), \( p(E_1 \cup E_2) = p(E_1) + p(E_2) \)

I.H.: Assume that \( P(k) \) holds for an arbitrary integer \( k \)

**Inductive step:** We will now show that \( P(k) \rightarrow P(k+1) \)

- Consider \( E = E_1 \cup E_2 \cup ... \cup E_k \cup E_{k+1} \)
- Let \( J = E_1 \cup E_2 \cup ... \cup E_k \), so \( E = J \cup E_{k+1} \)
- \( p(E) = p(J \cup E_{k+1}) \) by definition of \( E \)
- \[ = p(J) + p(E_{k+1}) \] by disjointedness
- \[ = p(E_1 \cup E_2 \cup ... \cup E_k) + p(E_{k+1}) \] by definition of \( J \)
- \[ = p(E_1) + p(E_2) + ... + p(E_k) + p(E_{k+1}) \] by I.H.

**Conclusion:** Since we have proved the base case and the inductive case, the claim holds by mathematical induction.