COMP S264F Unit 7: Discrete Probability

Dr. Keith Lee

School of Science and Technology

The Open University of Hong Kong

Overview

- Basics:
  - experiment
  - outcomes
  - sample space
  - event
- Probability of an event
- Puzzle 1: Envelopes game
- Combinations of events:
  - complement
  - union
  - intersection
- Puzzle 2: Dividing jobs between two cooks
- Probabilistic method: Set coloring problem

Discrete Probability

Applications in computer science:

- average case analysis of algorithms: Consider two algorithms for sorting n numbers. Suppose they use the same number of steps in the worst case. How can we further contrast their performance?
  Another example is network protocol.
- probabilistic algorithms: algorithms that make use random numbers (or coin flips); they may give wrong answer with some probability, but they are usually simple & quick. Example: testing whether an n-bit number is prime.
- data structures: hashing functions

Basic Definitions

- An *experiment* is a procedure that yields one of a given set of possible outcomes.
  E.g., roll a dice
- The set of possible outcomes is called the ***sample space***.
  E.g., possible outcomes: {1, 2, 3, 4, 5, 6}
- An event is a subset of the sample space.
  E.g., Large = {4, 5, 6}

Let us start with a simple assumption:
Experiment is based on a finite sample space of *equally likely* outcomes.

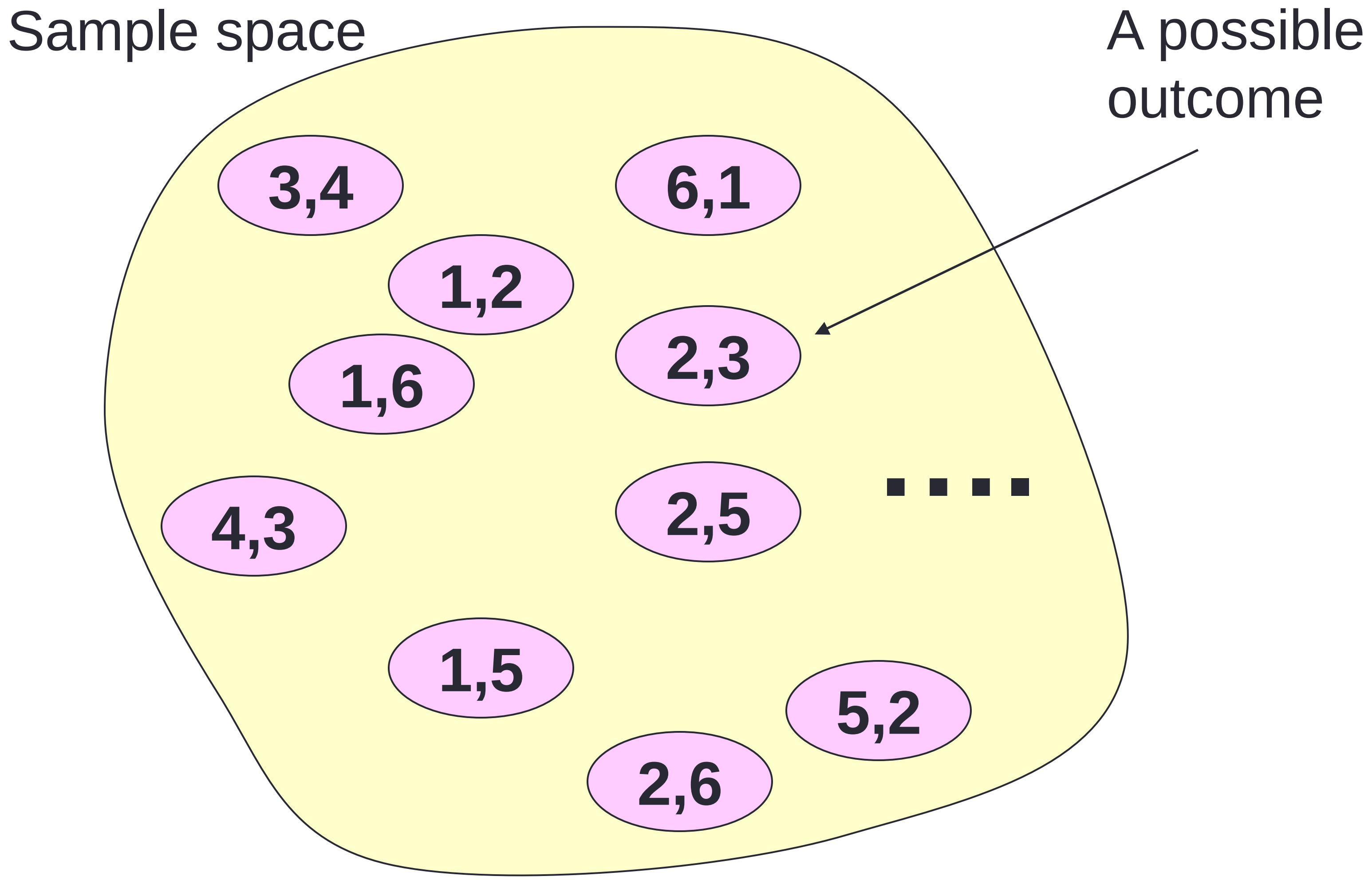
Probability

With respect to a sample space **S**, the ***probability*** of an event **E** (which is a subset of **S**), denoted by *p*(**E**), is **|E| / |S|**.

Example: What is the probability that when a dice is rolled, the number on the dice is "large"?
Answer: 3/6 = 1/2

Example: What is the probability that when two dice are rolled, the sum is 7?
Answer: No. of possible outcomes in the sample space = 36
No. of possible outcomes in the event = 6

Probability = 6/36 = 1/6


Sample space
A possible outcome
3,4
6,1
1,2
2,3
1,6
4,3
2,5
....
1,5
5,2
2,6

Example

Which is more likely (has a bigger probability),

rolling a total of 8 when two dice are rolled, or

rolling a total of 8 when three dice are rolled?

Probability =

Probability =

Lottery

Winner = correctly choose 6 numbers out of the numbers 1, 2, …, 47

What is the probability of being a winner?

|sample space| = C(47, 6)

|winner event| = 1

Answer = $\dfrac{1}{C(47,6)} = \dfrac{1}{10,737,573}$

Pennsylvania Super-lottery

- With one ticket, you can choose 7 numbers out of the first 80 numbers.
- Winning ticket = the 7 numbers chosen are among the 11 numbers selected by the lottery commission

- |sample space| = ?

- |winner event| = ?

- Winning probability = ?

Pennsylvania Super-lottery

- With one ticket, you can choose 7 numbers out of the first 80 numbers.
- Winning ticket = the 7 numbers chosen are among the 11 numbers selected by the lottery commission
- |sample space| = C(80, 11)
- |winner event| = C(80-7, 4)
- Winning probability

$$= \frac{C(73,4)}{C(80,11)} = \frac{1,088,430}{10,477,677,064,400}$$

NB. The **experiment** is the random process of selecting 11 numbers.

An **outcome** consists of 11 numbers (ordering is not important).

With respect to a ticket, the winning event is the set of all outcomes that contain the 7 numbers of the ticket.

Pennsylvania Super-lottery revisited

- The lottery commission has selected 11 numbers in advance.
- A lottery ticket contains 7 numbers which are chosen randomly by the computer.
- You win if the numbers in your ticket all appear in the 11 numbers chosen by the lottery commission.

- What is the experiment?
- What is an outcome?
- |sample space| =
- |winning event| =

Puzzle 1: Envelopes Game

- Keith puts a 1000-dollar bill in one of 3 envelopes and let you choose an envelope.

- Afterward, Keith opens one of his two envelopes that does not contain the 1000-dollar bill, and he asks you whether you would like to swap your envelope with the unopened one (a swap costs 5 dollars).
- **Question:** Should you pay for the “swap”? Does a “swap” really matter?

Let’s make a deal

- Keith has put a 1000-dollar bill in one of 3 envelopes and let you choose an envelope.
- If you are now asked whether the bill is with you or Keith, which would you pick?
- Obviously, ______________

- The probability that the 1000-dollar bill is with you is 1/3.
- The probability that the 1000-dollar bill is still with Keith is 2/3.

Two variations of the game

Variation 1

- Keith puts his two envelopes into a big envelope.
- Now you are asked again to choose your envelope or Keith's big envelope.
- What would be your choice?

- Prob [Your envelope contains \$1000] = 1/3
- Prob [Keith's big envelope contains \$1000] = 2/3

Variation 2

- Keith opens one of his two envelopes that doesn't contain the bill.
- You are asked again to choose an envelope.
- What would be your choice?

- Prob [Your envelope contains \$1000] = 1/3
- Prob [Keith's unopened envelope contains \$1000] = ???

A detailed study of Variation 2

- Let **x** denote the envelope containing $1000. Let **y** and **z** be the other envelopes.
- You choose an envelope randomly.

<u>3 possible outcomes</u>:

- You pick **x**: $1000 is with you.
- You pick **y**:
  - **x** and **z** are left to Keith.
  - Keith opens **z** and $1000 is in the only envelope with Keith.
- You pick **z**:
  - **x** and **y** are left to Keith.
  - Keith opens **y** and $1000 is in the only envelope with Keith.

Combinations of Events

- Let E be an event with respect to a sample space S. I.e., $E \subseteq S$.

- Let $\bar{E}$ be the complementary event of E, i.e., S - E.

- The probability of $\bar{E}$ is $\dfrac{|S - E|}{|S|}$

$$= \frac{|S|}{|S|} - \frac{|E|}{|S|}$$

$$= 1 - p(E).$$

Example 1

A sequence of 10 bits is generated randomly. What is the probability that at least one of these bits is 0?

- Let S be the sample space of generating 10 random bits. $|S| = 2^{10} = 1024$.
- Let E be the event that at least one of the 10 bits is 0. $|E| = ?$
- $\overline{E}$ is the event that all 10 bits are 1's. $|\overline{E}| =$
- $p(E) = 1 - p(\overline{E}) = 1 - \frac{1}{1024} = \frac{1023}{1024}$

Example 2

In a group of 5 randomly chosen people, the probability that at least two people were born in the same month is greater than 1/2.

True or False?

- What is an outcome?

Example 2

In a group of 5 randomly chosen people, the probability that at least two people were born in the same month is greater than 1/2.

- The underlying experiment is choosing 5 people randomly; an outcome refers to the months they were born.

- $|\text{sample space}| = 12\times12\times12\times12\times12 = 12^5 = 248{,}832$
- Let E be the event that *all five people were born in different months*.
- $|E| = 12\times11\times10\times9\times8 = 95{,}040$
- $p(E) = 95{,}040 / 248{,}832 = 0.38194444\ldots$
- The probability that at least two people were born in the same month is $1 - p(E) > 1/2$.

Union and Intersection

- Let E_1 and E_2 be two events in a sample space S.
- Is the union of E_1 and E_2 an event?
- What about their intersection?

- **Theorem:** $p(E_1 \cup E_2) = p(E_1) + p(E_2) - p(E_1 \cap E_2)$
- **Corollary:** $\mathbf{p(E_1 \cup E_2) \leq p(E_1) + p(E_2)}$

- E.g., What is the probability that an integer selected at random from {1, 2, …, 100} is divisible by 2 or 5?

Puzzle 2: Dividing jobs between two cooks

- A Chinese restaurant offers 88 different dinner sets (meal plans), each set comprising 8 dishes, chosen from 100 possible dishes.

- The restaurant has two famous cooks, named Blue and Red. Each cook is responsible for a subset of the 100 dishes.

- Is there a way to divide the 100 dishes among the cooks so that both of them are involved in every dinner set?

Example

- Assume that 100 dishes are labeled with D_1, D_2, …, D_{100}.

- Set dinner $A_1 = \{D_5, D_{12}, D_{41}, D_{44}, D_{45}, D_{67}, D_{78}, D_{91}\}$
- Set dinner $A_2 = \{D_2, D_5, D_{12}, D_{14}, D_{15}, D_{27}, D_{29}, D_{91}\}$
- …
- Set dinner $A_{88} = \{D_{12}, D_{51}, D_{56}, D_{66}, D_{69}, D_{70}, D_{78}, D_{88}\}$

- Blue: D_1, D_2, D_5, …, D_{66}, …
- Red: D_3, D_4, D_6, D_{12}, …

A Set Coloring Problem: an example of probabilistic method

Input: Let **A** be a set with n elements.

Let A_1, A_2, …, A_m be distinct subsets of **A**, each containing w elements (we assume that $m < 2^{w-1}$).

Valid coloring: Color each element of **A** red or blue such that each A_i contains at least one red element and one blue element.

Question: Does there exist a valid coloring of A?

In our cook puzzle, n = 100, m = 88, w = 8.

Probabilistic method

Three approaches to proving something (say, a valid coloring) to exist.

- Devise an algorithm to find a valid coloring.
- Proof by contradiction: Assume that a valid coloring doesn't exist. … A contradiction occurs.
- Show that a **random** coloring of A has probability > 0 being a valid coloring. Then there exists a valid coloring.

Valid Coloring always exists!

- Suppose we color each element of **A** at random (say, flip a coin for each element).
- sample space = { (blue, blue, red, blue, …, red), … }
- |sample space| = 2^n.

- Consider a particular subset A_i .
- Let E_i be the event that **all** elements of A_i are colored red.
- Then $|E_i| = 2^{n-|A_i|} = 2^{n-w}$.
- $p(E_i) = \dfrac{|E_i|}{|\text{sample space}|} = \dfrac{2^{n-w}}{2^n} = \dfrac{1}{2^w}$.

- Let **E** be the event that among the sets $A_1, A_2, \ldots, A_m$, there is at least *one subset* with all elements colored red.

- **E** can be rephrased as the event that
  - A_1 has all elements colored red, or
  - A_2 has all elements colored red, or
  - …, or
  - A_m has all elements colored red.
- $$\begin{aligned} p(\mathbf{E}) &= p(E_1 \cup E_2 \cup E_3 \cup \ldots \cup E_m) \\ &\le p(E_1) + p(E_2 \cup E_3 \cup \ldots \cup E_m) \\ &\le p(E_1) + p(E_2) + p(E_3 \cup \ldots \cup E_m) \\ &\le \ldots \le p(E_1) + p(E_2) + p(E_3) + \ldots + p(E_m) = \frac{m}{2^w} . \end{aligned}$$

- Let **F** be the event that there is at least *one subset* with all elements colored blue.

- By symmetry, $p(\mathbf{F}) \le \frac{m}{2^w}$.

Conclusion

$E \cup F$ refers to the event that there is *<u>one subset</u>* with elements colored all red or all blue, i.e., we get an <u>invalid</u> coloring.

$p(E \cup F) \leq p(E) + p(F) \leq \frac{2m}{2^w} = \frac{m}{2^{w-1}}$ **< 1** (because we assume

that $m < 2^{w-1}$).

In other words,

- the probability of getting an invalid coloring is < 1,
- the probability of getting a valid coloring is > 0, and
- there must exist at least one coloring in the sample space that is a valid coloring.