COMP S264F Unit 5: Basics of Counting

Dr. Keith Lee

School of Science and Technology

The Open University of Hong Kong

Overview

- Sum rule
- Product rule

- Principle of Inclusion-Exclusion

- Pigeonhole principle
- Generalized pigeonhole principle

- Puzzle: Increasing / decreasing subsequence

Counting Problems

- Counting problems arise on many occasions in Computer Science.

Examples:

- A computer password consists of 6, 7, or 8 characters, each must be a digit or a letter.
  A password must also contain at least one digit.
  How many passwords are there?

- Given a complicated nested loop involving conditionals, count the number of iterations.

Basics of Counting: Sum and Product

- A student can choose an elective course (e.g., general education course) from two schools.
- These schools offer 3 and 5 courses, respectively.

How many possible courses are there to choose from?

Answer: 3+5 = 8.

NB. This is true only if the two schools do not offer a joint course. Otherwise, 3+5 is over-counted!

The Sum Rule

Suppose that a task can be done either in one of x_1 ways or in one of x_2 ways, where
none of the x_1 ways is the same as any one of the x_2 ways (i.e., the two sets of ways are disjoint).

Then there are $x_1 + x_2$ ways to do the task.

The Product Rule

How many different license plates are available if each contains two letters followed by 4 digits?

Answer: $26 \times 26 \times 10 \times 10 \times 10 \times 10 = 6{,}760{,}000$

Suppose that a procedure can be broken in n steps
$T_1, T_2, \ldots, T_n$, and
T_i can be done in y_i ways after $T_1, T_2, \ldots, T_{i-1}$ have been done (in whatever ways).

Then there are $y_1 \times y_2 \times \ldots \times y_n$ ways to do the procedure.

Examples

How many functions are there from a set A with m elements to a set B with n elements?

Solution:

- Let $A = \{a_1, a_2, \ldots, a_m\}$.
- The following procedure generates a function f from A to B:

  m steps: $T_1, T_2, \ldots, T_m$

  where T_i assigns an element of B to a_i.
- Each step can be done in n ways.
- There are n^m ways to generate f.

How many one-to-one functions are there?

An old example (Unit 3 Slide 35)

- Let A be a set of n elements.
- What is the cardinality of the set

$$\{ (X, Y) \mid X \subseteq A,\ Y \subseteq A,\ X \cap Y = \varnothing \}?$$

Solution:

- Let $A = \{a_1, a_2, \ldots, a_n\}$.
- Consider the following procedure for choosing an ordered pair (X, Y):

  n steps: $T_1, T_2, \ldots, T_n$

  where T_i puts a_i into one of the followings:
  - X
  - Y
  - none

- Each T_i can be done in 3 different ways.
- The procedure can be done in 3^n ways.

Using both Sum rule & Produce rule

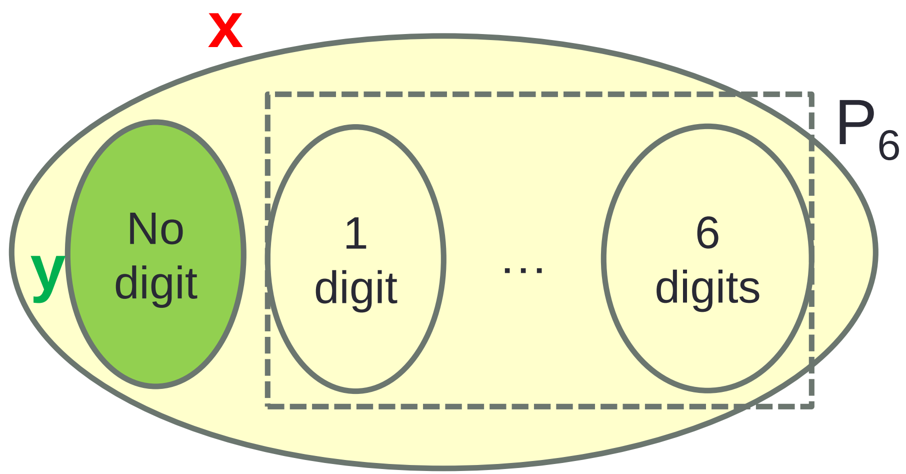
Example: A computer password consists of 6, 7, 8 characters, each must be a digit or a letter. A password must also contain at least one digit. How many passwords are there?

Solution:

- Let P be the total number of such passwords.
- Let P_6, P_7, and P_8 denote the number of passwords of length 6, 7, and 8, respectively.

- By the sum rule, $P = P_6 + P_7 + P_8$.

What is P_6 ?



- 62 possible characters:
  - 26 uppercase letters
  - 26 lowercase letters
  - 10 digits
- The number of all possible 6-character strings is 62^6, which is denoted by **x**.
- The number of 6-character strings containing no digit is 52^6, which is denoted by **y**.
- P_6 = the number of 6-character strings containing at least one digit.
- By the sum rule, **x** = P_6 + **y**.
- Therefore, $P_6 = 62^6 - 52^6 = 37{,}029{,}625{,}920$.

- $P_6 = 62^6 - 52^6 = 37{,}029{,}625{,}920$.

Similarly,

- $P_7 = 62^7 - 52^7 = 2{,}493{,}542{,}903{,}680$.
- $P_8 = 62^8 - 52^8 = 164{,}880{,}377{,}053{,}440$.

Therefore,

- $P = P_6 + P_7 + P_8$

  $= 167{,}410{,}949{,}583{,}040$.

How many bit strings of length 10 contain 5 consecutive 0's?

0 1 0 0 0 0 0 1 1 0

- For any integer $i \leq 10$, let P_i be the number of length-10 strings containing 5 consecutive 0's starting from the position i.
- Answer = $P_1 + P_2 + P_3 + P_4 + P_5 + P_6$

How many bit strings of length 10 contain 5 consecutive 0's?

- Answer = $P_1 + P_2 + P_3 + P_4 + P_5 + P_6$
  where P_i is the number of strings of length 10 containing 5 consecutive 0's with the first occurrence of 5 consecutive 0's starting from position i.

- $P_1 = 2^5 = 32$

0 0 0 0 0 1 1 0 0 0

- P_2 = ?

1 0 0 0 0 0 1 1 0 0

- P_3 = ?

0 1 0 0 0 0 0 1 1 0

- Answer = ?

More counting examples

Calculate the number of length-7 strings over the alphabet {a, b} that begin with an a, and contain at least one b.

Solution:

- The number of choices for the 1st character is 1.

- b can only occur after the 1st character.
- The number of choices for the other 6 characters
  = (number of all possible length-6 strings)
  - (number of all length-6 strings without any b)
  = 2^6 - 1
  = 64 - 1 = 63

- Answer = 1 $\times$ 63 = 63

More counting examples (cont')

Calculate the number of length-6 strings over the alphabet {a, b, c} that begin with an a or b, and contain at least one c.

Solution:

- The number of choices for the 1st character is 2.

- c can only occur after the 1st character.
- The number of choices for the other 5 characters
  = (number of all possible length-5 strings)
  - (number of all length-5 strings without any c)
  = $3^5 - 2^5$
  = 243 - 32 = 211

- Answer = $2 \times 211 = 422$

Over Counting

- How many bit strings of length 8
  - start with 1; or
  - end with 00 ?

- Number of bit strings starting with 1 = 2^7 = 128.
- Number of bit strings ending with 00 = 2^6 = 64.
- Answer = 128 + 64?

- No. The string **1**111000**00** is double counted!

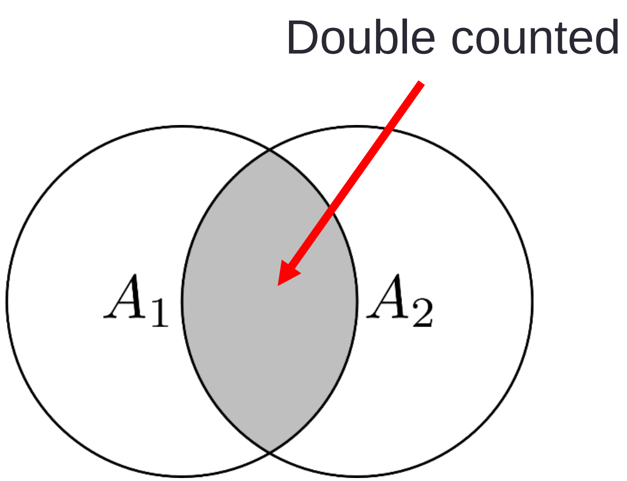
Principle of Inclusion-Exclusion

- Let A_1 be the set of ways a task T_1 can be done, and let A_2 be the set of ways a task T_2 can be done.
- Note that A_1 may overlap A_2, i.e., there are some ways that satisfy both T_1 and T_2.
- Then, the number of ways to do T_1 or T_2 is

$$|A_1 \cup A_2| \;=\; |A_1| + |A_2| - |A_1 \cap A_2|$$

Example:

- Number of bit strings of length 8 that start with 1 ***and*** end with 00 $= 2^5 = 32$
- Number of bit strings of length 8 that start with 1 ***or*** end with 00 $= 128 + 64 - 32 = 160$



Another example

- How many bit strings of length 10 contain 5 consecutive 0's or 5 consecutive 1's?

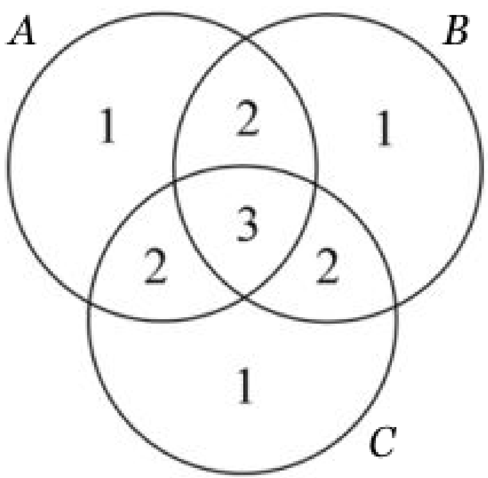
Solution:

- 5 consecutive 0's: 112
- 5 consecutive 1's: 112
- 5 consecutive 0's and 5 consecutive 1's: 2
- Answer = 112 + 112 – 2 = 222

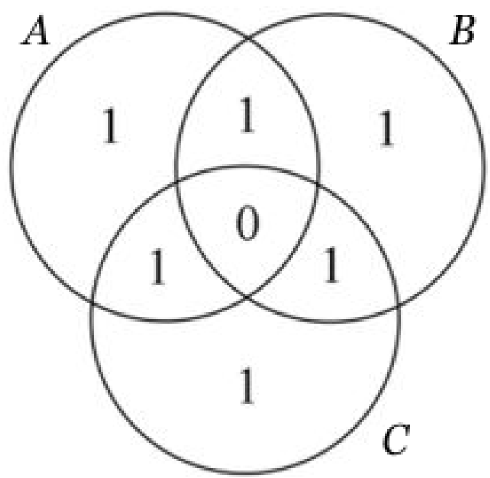
Principle of Inclusion-Exclusion for 3 sets

Let A, B, C be three sets. Then,

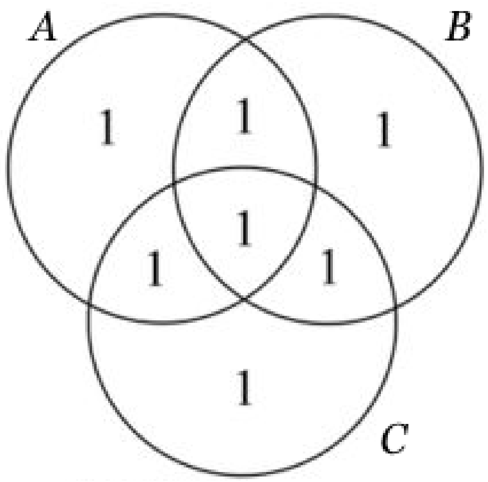
$$\begin{aligned} |A \cup B \cup C| &= |A| + |B| + |C| \\ &\quad - |A \cap B| - |A \cap C| - |B \cap C| \\ &\quad + |A \cap B \cap C| \end{aligned}$$



$|A| + |B| + |C|$



$|A| + |B| + |C|$
$- |A \cap B| - |A \cap C| - |B \cap C|$



$|A| + |B| + |C|$
$- |A \cap B| - |A \cap C| - |B \cap C|$
$+ |A \cap B \cap C|$

NB. Number in a subset
= Number of times that subset is counted.

Example

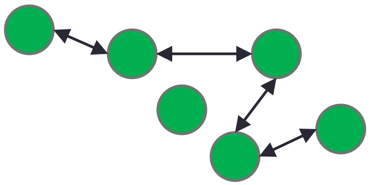
- In a class, 15 students study C programming;
  20 students study Java; 13 students study Python;
  5 students study both C and Java;
  7 students study both C and Python;
  4 students study both Java and Python; and
  no students study all the three programming languages.
- What is the number of students in the class?

Solution:

- Let C, J, P be the set of students on C, Java, Python, respectively.
- The number of students in the class is

$$\begin{aligned} |C \cup J \cup P| &= |C| + |J| + |P| - |C \cap J| - |C \cap P| - |J \cap P| \\ &\quad + |C \cap J \cap P| \\ &= 15 + 20 + 13 - 5 - 7 - 4 + 0 \\ &= 32\ . \end{aligned}$$

Friends or not



- There are 6 people inside the elevator. Suppose that if a person A knows a person B, then B also knows A.

True or false?

- There are at least 2 people who know each other or don't know each other. True

- There are at least 3 people who know each other or don't know each other.

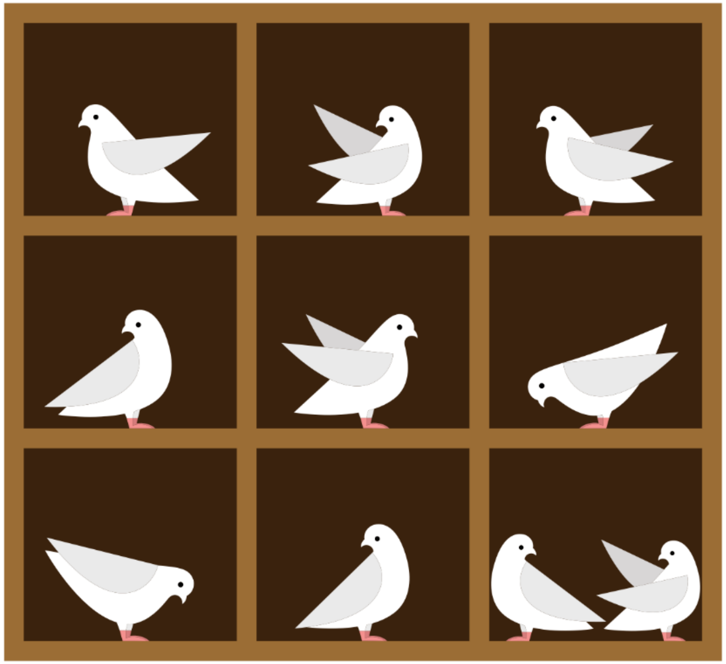
- There are at least 4 people who know each other or don't know each other.

The Pigeonhole Principle

Theorem: If k+1 or more objects are placed into k boxes, then there is at least one box containing *two or more* of the objects.

Example:

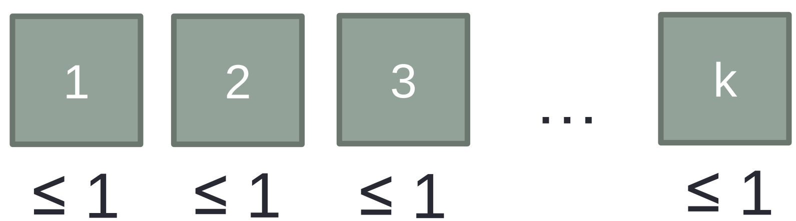
- 10 pigeons
  in 9 pigeonholes



The Pigeonhole Principle: Proof

- Consider any way of placing k+1 or more objects into k boxes. (I.e., the boxes contain at least k+1 objects in total.)

- Suppose, for the sake of contradiction, that no box contains two or more objects.



- Let us sum up the objects in the k boxes.
  The sum is at most k.

- A contradiction occurs.

Generalized pigeonhole principle

If n objects are placed into k boxes, then there is at least one box containing at least $\left\lceil \frac{n}{k} \right\rceil$ objects.

The smallest integer $\geq \frac{n}{k}$

Proof: Again, by contradiction.

- Suppose there is not any box containing at least $\left\lceil \frac{n}{k} \right\rceil$ objects.
- I.e., the number of objects in every box is $< \left\lceil \frac{n}{k} \right\rceil$, or equivalently, $\leq \left\lceil \frac{n}{k} \right\rceil - 1$.
- Thus, the total number of objects in the k boxes is

$$\leq k \cdot \left(\left\lceil \frac{n}{k} \right\rceil - 1 \right) < k \cdot \left(\frac{n}{k} + 1 - 1 \right) = n \ .$$

- A contradiction occurs.

Example 1

Find the minimum number of people needed such that three of them were born on the same day of the week?

Solution:

- There are 7 days of week, namely, Monday, …, Sunday.
- We need to find the minimum number n of people such that by generalized pigeonhole principle, one of these days of week has at least $\left\lceil \frac{n}{7} \right\rceil = 3$ people.
- Therefore, $n = 7 \times 2 + 1 = 15$.

To check that the minimum n is 15, consider the case with 14 people. It is possible that each day of week has exactly 2 people.

Example 2

Find the minimum number of people needed such that four of them were born in the same month?

Solution:

- There are 12 months.
- We need to find the minimum number *n* of people such that by generalized pigeonhole principle, one of these months has at least $\left\lceil \frac{n}{12} \right\rceil = 4$ people.
- Therefore, *n* = 12×3 + 1 = 37.

To check that the minimum *n* is 37, consider the case with 36 people. It is possible that each month has exactly 3 people.

Example 3

Why does any set of 10 non-empty strings over the alphabet {a, b, c} contain two different strings with the same starting character and the same ending character?

Solution:

- Given any non-empty string, let x and y be its starting and ending characters.
- (x, y) has $3 \times 3 = 9$ distinct values.
- By pigeonhole principle, in any set of 10 non-empty strings, there must be two different strings with the same (x, y) values, i.e., they have the same starting characters and the same ending characters.

Example 4

Find the minimum number of non-empty strings over {a, b, c, d} such that three of them have the same starting character and the same ending character.

Solution:

- Given any non-empty string, let x and y be its starting and ending characters.
- (x, y) has $4 \times 4 = 16$ distinct values.
- We need to find the minimum number *n* of non-empty strings such that by generalized pigeonhole principle, one of these (x, y) pair values has at least $\left\lceil \frac{n}{16} \right\rceil = 3$ strings.
- Therefore, $n = 16 \times 2 + 1 = 33$.

Friends or not: Proof

- *Assumption:* If a person A knows a person B, then B also knows A. In this case, we say that A and B are friends.

- *Trivial fact:* For any two persons, either they know each other or they don't know each other.

True or false?

- Out of any 6 people, there are at least 4 people who know each other or don't know each other.

Friends or not: Proof

Claim: Out of any 6 people, there are at least 3 people who know each other or don't know each other.

Proof.

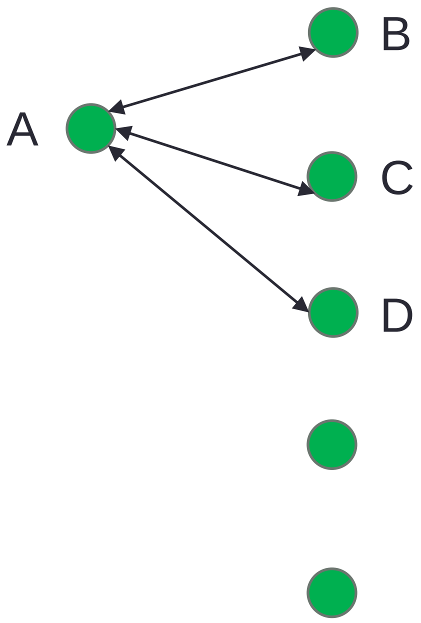
- Let **A** be one of the six people.
- Of the other five people, there are
  - either three or more who know **A** (are friends of **A**),
  - or three or more who don't know **A** (are not friends of **A**).

Why?

- The five people can be classified into 2 groups, i.e., friends of A and not friends of A.
- By the generalized pigeonhole principle,

  one group has a size $\geq \left\lceil \frac{5}{2} \right\rceil = 3$.

Friends or not: Proof (cont')



Let us first consider the case when **A** has 3 or more friends.

Name 3 of them as B, C, and D.

- If two of B, C, and D are friends, then these two and A form a group who know each other.
- If any two of B, C, and D are not friends, then B, C, D form a group who don't know each other.

The other case when **A** does not have 3 or more friends is symmetric. Fill in the details yourself.

Puzzle: Increasing/decreasing subsequence

Theorem: Let S be a sequence of n^2+1 distinct numbers. Then S contains a *subsequence* of length $n+1$ (not necessarily consecutive) that is either strictly *increasing* or strictly *decreasing*.

Example: $n = 3$, $S = (45, 2, 39, 111, 32, 4, 99, 1, 23, 0)$

- Two decreasing subsequences of length 4:

1. 111, 99, 1, 0
2. 111, 32, 4, 1

- Any increasing subsequence of length 4?

Puzzle: Proof

Let $S = (a_1, a_2, \ldots, a_{n^2+1})$ be a sequence of distinct numbers.

For each a_k, define

- I_k = the length of the longest ↑ (increasing) subsequence starting from a_k;
- D_k = the length of the longest ↓ (decreasing) subsequence starting from a_k.

E.g., S = (45, 2, 39, 111, 32, 4, 99, 1, 23, 0)

- For k = 5, a_k = 32; I_5 = 2 and D_5 = 4.
  ↑ : (32, 99)
  ↓ : (32, 4, 1, 0)

- For k = 9, a_k = 23; I_9 = 1 and D_9 = 2.
  ↑ : (23)
  ↓ : (23, 0)

- Suppose, for the sake of contradiction, that every ↑ or ↓ subsequence of S is of length at most *n*.
- Then, for any k, $I_k \leq n$ and $D_k \leq n$.

- Therefore, (I_k, D_k) can have at most n^2 distinct values, namely, (1, 1), (1, 2), …, (n, n).
- The index k ranges from 1 to n^2+1.

- By pigeonhole principle, among the n^2+1 pairs

$$(I_1, D_1), (I_2, D_2), \ldots, (I_{n^2+1}, D_{n^2+1})$$

$\exists\ k_1, k_2$ with $k_1 < k_2$
such that (I_{k_1}, D_{k_1}) and (I_{k_2}, D_{k_2}) have the same value (i.e., $I_{k_1} = I_{k_2}$ and $D_{k_1} = D_{k_2}$).

Two cases to consider:

- Case 1: $a_{k_1} < a_{k_2}$.
  We can form an ↑ subsequence of length I_{k_2}+1 starting from a_{k_1}, followed by a_{k_2} …

$$S = a_1\ a_2 \ldots a_{k_1} \ldots \underbrace{a_{k_2} \ldots\ldots\ldots\ldots\ldots\ldots\ldots a_{n^2+1}}_{\uparrow \text{ subsequence of length } I_{k_2}}$$

By the definition of I_{k_1}, the longest ↑ subsequence starting from a_{k_1} is of length $I_{k_1} = I_{k_2}$.

A contradiction occurs.

- Case 2: $a_{k_1} > a_{k_2}$. The argument is similar to Case 1.

We can form a ↓ subsequence of length $D_{k_2}+1$ starting from a_{k_1}, followed by a_{k_2} …

$$S = a_1\ a_2 \ldots a_{k_1} \ldots \underbrace{a_{k_2} \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ a_{n^2+1}}_{\downarrow \text{ subsequence of length } D_{k_2}}$$

By the definition of D_{k_1}, the longest ↓ subsequence starting from a_{k_1} is of length $D_{k_1} = D_{k_2}$.

Again, a contradiction occurs.

<u>In conclusion</u>, S must contain an ↑ or ↓ subsequence of length *n+1*.