50.034 - Introduction to Probability and Statistics

Week 1 – Cohort Class

January–May Term, 2019

SINGAPORE UNIVERSITY OF TECHNOLOGY AND DESIGN

Outline of Cohort Class

- Counting methods
- Multiplication Rule and Tuples
- Permutations and Combinations
- Class Activity: Inclusion-exclusion principle

Motivation

Recap of definitions covered this week:

- An experiment is any process (real or hypothetical), in which the possible outcomes can be identified ahead of time.
- The sample space of an experiment, usually denoted by Ω, is the set of **ALL** possible outcomes of that experiment.
- An event is a subset of outcomes contained in the sample space Ω.

Suppose a sample space Ω has N outcomes.
For any event A, let $N(A)$ be the number of outcomes in A.

When all outcomes of an experiment are equally likely, the task of computing $\Pr(A)$ reduces to counting:

$$\Pr(A) = \frac{N(A)}{N}$$

In today's class, we study methods that help us count $N(A)$ and N.

How many ways?

With two shirts (S1, S2) and three pants (P1, P2, P3), how many ways can we create an outfit?

Answer: $2 \times 3 = 6$ ways.

Multiplication rule

Suppose we want to choose a pair of objects from some given collection of objects, such that the first object satisfies property P_1, and the second object satisfies property P_2.

If the first object can be selected in n_1 ways, and if for each of these n_1 possible selections, the second object can be selected in n_2 ways, then the total number of possible pairs is $n_1 n_2$.

Example: A class has 12 boys and 18 girls. The teacher selects a pair consisting of 1 boy and 1 girl. She can do this in any of $12 \times 18 = 216$ ways.

Tuple

A sequence with k entries is called a k-tuple.

- e.g. a pair is a 2-tuple and a triple is a 3-tuple.

Multiplication rule for k-tuples:
Suppose that

- the 1st entry can be selected in n_1 ways;
- the 2nd entry can be selected in n_2 ways;
- $\ldots$
- the k-th entry can be selected in n_k ways;

such that all possible selections for each entry could be selected, regardless of which specific selections have been made in the other entries.

Then there are $n_1 n_2 \cdots n_k$ possible k-tuples.

Example 1

Suppose a home remodeling job involves purchasing appliances, plumbing and setting up electrical wires. There are 5 dealers for appliances, 12 plumbing contractors, and 9 electrical workers in the area to choose from.

How many ways are there to form a home remodeling team?

Answer:
We can think of the team as a 3-tuple:

$$(\text{appliance dealer}, \text{plumbing contractor}, \text{electrical worker}).$$

By the multiplication rule for 3-tuples, we can form this team in $5 \times 12 \times 9 = 540$ ways.

Permutation

Consider a collection of n **distinct** objects.

Suppose that an experiment consists of selecting k objects one at a time, without replacement, i.e. objects that are already selected cannot be selected subsequently.

Each possible outcome of the experiment is a k-tuple, which is called a permutation of n objects taken k at a time, or more simply, a permutation of size k.

- When selecting distinguishable objects, **the order matters!** If we select object A first, followed by object B, that is not the same as selecting object B first, followed by object A.

Question: How many such permutations are there?

We usually denote the number of such permutations by $P_{k,n}$.

$P_{k,n}$ can be determined from the multiplication rule for k-tuples.

Example 2

In a group of 10 students, how many ways are there to select a list of 5 students, say for the five positions of "president", "vice-president", "secretary", "treasurer", and "logistics head"?

Answer: $P_{5,10} = 10 \times 9 \times 8 \times 7 \times 6 = 30240$.

There are 10 ways to select the 1st student (as president).

There are 9 ways to select the 2nd student (as vice-president) from the remaining 9 students.

- Regardless of who the 1st student is, there are always 9 ways to select the 2nd student.

There are 8 ways to select the 3rd student (as secretary) from the remaining 8 students.

- Again, regardless of who the 1st and 2nd students are, there are always 8 ways to select the 3rd student.

So on and so forth.

Example 3

There are 10 students: A, B, C, D, E, F, G, H, I and J.

Suppose we want to select a list of 3 students among these 10 students, for three different positions.

1. How many ways are there to select a list of 3 students?
2. How many ways are there to select a list of 3 students, such that student A is the middle student of the list?
3. If every student is equally likely to be selected, what is the probability that student A is the middle student of the list?

Example 3 - Solution

1. The number of possible lists is $P_{3,10} = 10 \times 9 \times 8 = 720$.

2. There are $P_{2,9} = 9 \times 8 = 72$ ways to select 2 students, such that the first among the two is in front of student A, and the second among the two is behind student A.

3. To find the probability, we need to find $\frac{N(A)}{N}$, where N is the number of ways to select a list of 3 students, and $N(A)$ is the number of ways to select a list of 3 students such that student A is the middle student.

Hence, the probability is $\frac{N(A)}{N} = \frac{72}{720} = \frac{1}{10}$.

Exercise 1 (10 mins)

There are 6 students: A, B, C, D, E and F.

Suppose we want to select a list of 4 students among these 6 students, for four different positions.

1. How many ways are there to select a list of 4 students?
2. How many ways are there to select a list of 4 students, such that student A is one of the middle two students of the list?
3. If every student is equally likely to be selected, what is the probability that student A is one of the middle two students of the list?

Exercise 1 - Solution

1. The number of possible lists is $P_{4,6} = 6 \times 5 \times 4 \times 3 = 360$.

2. Student A must either be the 2nd or 3rd student of the list.
If student A is the 2nd student of the list, then:

- There are $P_{3,5} = 5 \times 4 \times 3 = 60$ ways to select 3 students, such that one is in front of A, and two are behind A.

If student A is the 3rd student of the list, then:

- There are $P_{3,5} = 5 \times 4 \times 3 = 60$ ways to select 3 students, such that two are in front of A, and one is behind A.

Thus, total number of ways is $60 + 60 = 120$.

3. To find the probability, we need to find $\frac{N(A)}{N}$, where N is the number of ways to select a list of 4 students, and $N(A)$ is the number of ways to select a list of 4 students such that student A is one of the middle two students.

Hence, the probability is $\frac{N(A)}{N} = \frac{120}{360} = \frac{1}{3}$.

Combination

Definition: An unordered subset of a set is called a combination.

The number of combinations of size k that can be formed from n objects is denoted by $C_{k,n}$ or $\binom{n}{k}$.

- When the notation $\binom{n}{k}$ is used, this number is called a binomial coefficient.
- $\binom{n}{k}$ is usually read as "n choose k".

Formula to calculate binomial coefficients:

$$\binom{n}{k} = \frac{P_{k,n}}{k!} = \frac{n!}{k!(n-k)!}$$

The number of combinations is the number of corresponding permutations disregarding the different outcomes due to order.

Example 4

How many ways are there to select a list of 5 students, for five different positions, from a group of 10 students?

Answer: $P_{5,10} = \dfrac{10!}{(10-5)!}$.

How many ways are there to select a list of 5 students, for five different positions, from a group of 5 students?

Answer: $P_{5,5} = 5!$.

How many ways are there to select 5 students from a group of 10 students, to form a student committee?

Answer: $C_{5,10} = \dfrac{10!}{5!(10-5)!}$.

Example 5

There are 6 students: A, B, C, D, E and F.

Suppose we want to select 3 students among these 10 students to form a student committee.

1. How many ways are there to select a committee of 3 students?
2. How many ways are there to select a committee of 3 students, such that student A is in the committee?
3. If every student is equally likely to be selected, what is the probability that student A is in the committee?

Example 5 - Solution

1. The number of possible ways is $\binom{10}{3} = \frac{P_{3,10}}{3!} = \frac{10\times9\times8}{1\times2\times3} = 120$.

2. If student A must be in the committee, then there are $\binom{9}{2} = \frac{P_{2,9}}{2!} = \frac{9\times8}{1\times2} = 36$ ways to choose the remaining two students in the committee.

3. To find the probability, we need to find $\frac{N(A)}{N}$, where N is the number of ways to select a committee of 3 students, and $N(A)$ is the number of ways to select a committee of 3 students such that student A is in the committee.

Hence, the probability is $\frac{N(A)}{N} = \frac{36}{120} = \frac{3}{10}$.

Example 6

A music library has 100 songs, 10 of which are by the Beatles.

The 100 songs are played in a random order.

What is the probability that the first Beatles song played is the fifth of all the songs?

Hint: In order for this event to happen, the first 4 songs played must not be by the Beatles (non-B) and the fifth song is by the Beatles (B).

Example 6 - Method 1

Let the event (first 4 songs non-B, 5th song B) be denoted by E.

The probability is $\Pr(E) = \frac{N(E)}{N}$, where N is the number of ways to select 5 songs, and $N(E)$ is the number of ways that the event E can happen.

$$N = P_{5,100} = 100 \times 99 \times 98 \times 97 \times 96.$$

The number of ways to select 4 non-B songs, followed by 1 B song is

$$N(E) = P_{4,90} \times P_{1,10} = 90 \times 89 \times 88 \times 87 \times 10.$$

Therefore, $\Pr(E) = \dfrac{90 \times 89 \times 88 \times 87 \times 10}{100 \times 99 \times 98 \times 97 \times 96}$.

Example 6 - Method 2

We could also calculate $\Pr(E)$ using combinations.
Every playlist can be represented by a 100-tuple, whose entries are either 'B', or 'non-B'.

- Since there are 10 B songs, there are $\binom{100}{10}$ different ways to choose which entries of the 100-tuple would be filled by 'B'.
- Once we have chosen 10 entries to be filled by 'B', the remaining entries must then be filled by 'non-B'.

If event E occurs, then the 100-tuple must look like:

$$(\text{non-B}, \text{non-B}, \text{non-B}, \text{non-B}, \text{B}, \dots).$$

Since there are a total of 10 B songs, the remaining 95 entries must contain exactly 9 B's.
There are $\binom{95}{9}$ ways to fill up these remaining 95 entries by 'B'.
Therefore, $\Pr(E) = \frac{\binom{95}{9}}{\binom{100}{10}}$.
(Can you see why this value equals $\frac{90\times 89\times 88\times 87\times 10}{100\times 99\times 98\times 97\times 96}$?)

How to think?

It takes a lot of practice to extract the "mathematical core" beneath a physical description, which may be covered by various kinds of fancy stories.

To count the number of outcomes satisfying a certain given condition, think of how you can reinterpret the condition in terms of **k-tuples**, **permutations**, and/or **combinations**, and use the **multiplication rule**.

Try to convince your friends that your counting method is correct. If you can do so, then you are probably correct.

Remember, there could be more than one method to count.

Exercise 2 (20 mins)

Twelve different books, consisting of 5 math books, 4 science books, and 3 history books are arranged in order on a bookshelf. Find the number of arrangements in each of the following cases:

1. There are no restriction.
2. The books of each type must be together.
3. The math books must be together.

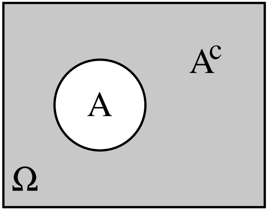
Exercise 2 - Solution

1. $12!$ (Permute all 12 books.)

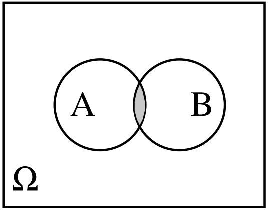
2. $5! \cdot 4! \cdot 3! \cdot 3!$ (Permute within each type of books: $5! \cdot 4! \cdot 3!$; then permute the three types: $3!$.)

3. $5! \times 8!$ (Permute the math books: $5!$; then permute the rest of the books and the group of math blocks, by considering the group of math books as one big "book": $8!$.)

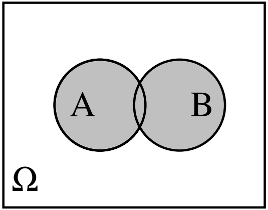
Recall Venn diagrams



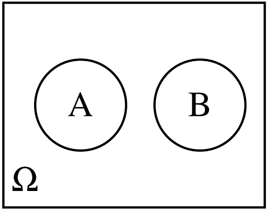
A^c shaded



$A \cap B$ shaded



$A \cup B$ shaded



Disjoint events

Exercise 3 (15 mins)

Of the students in a class, 60% are geniuses, 70% love chocolate, and 25% fall into neither group. Determine the probability that a randomly selected student is a genius or a chocolate lover but not both.

Exercise 3 - Solution

Let G be the event that the student selected is a genius.
Let C be the event that the student selected is a chocolate lover.
Given info: $\Pr(G) = 0.6$, $\Pr(C) = 0.7$, and $\Pr(G^c \bigcap C^c) = 0.25$.

We have

$$\Pr(G \cup C) = \Pr(G) + \Pr(C) - \Pr(G \cap C) = 1 - \Pr(G^c \cap C^c).$$

This means that $0.6 + 0.7 - \Pr(G \cap C) = 1 - 0.25$, or equivalently, $\Pr(G \cap C) = 0.55$.

Since $\Pr(G) = \Pr(G \bigcap C^c) + \Pr(G \bigcap C)$,

$$\Pr(G \cap C^c) = \Pr(G) - \Pr(G \cap C) = 0.6 - 0.55 = 0.05.$$

Similarly,

$$\Pr(C \cap G^c) = \Pr(C) - \Pr(C \cap G) = 0.7 - 0.55 = 0.15.$$

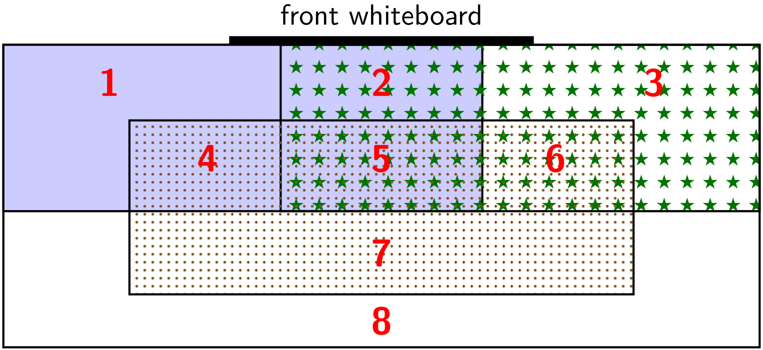
Exercise 3 - Solution

Therefore,

$$
\begin{aligned}
&\Pr((G \cap C^c) \cup (C \cap G^c)) \\
&= \Pr(G \cap C^c) + \Pr(C \cap G^c) - \Pr((G \cap C^c) \cap (C \cap G^c)) \\
&= \Pr(G \cap C^c) + \Pr(C \cap G^c) = 0.05 + 0.15 = 0.2.
\end{aligned}
$$

Class Activity (15 mins)

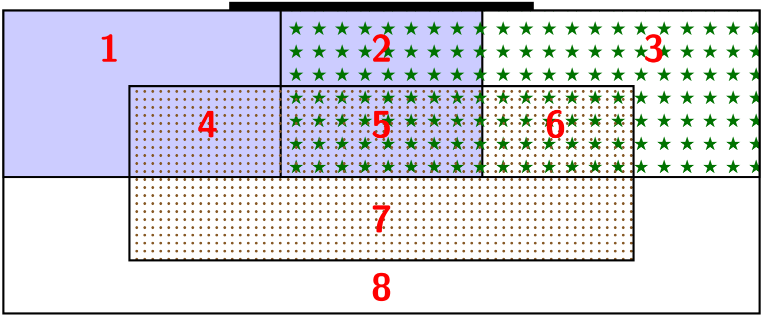
It's time to get up!



There are a total of 8 zones. Move over to your zone, according to the following criteria:

- **Blue Zones:** Your birthday falls within 1 January–30 June.
- **Star Zones:** You have been to Macdonalds in the past month.
- **Dotted Zones:** You have an elder brother or sister.

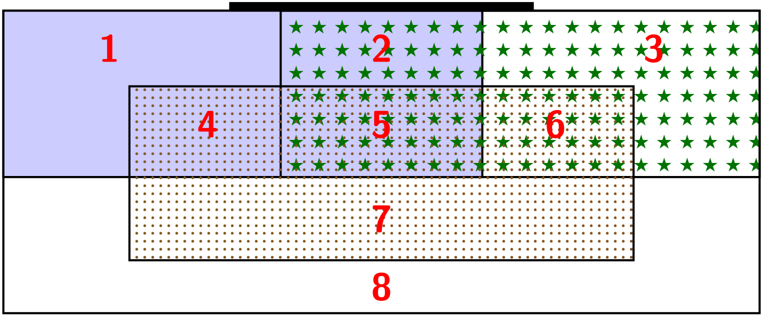
Class Activity (15 mins)



Question: How many of you are born in the first half of the year, **or** have been to Macdonalds in the past month, **or** have an elder brother or sister? (i.e. How many of you are in zones 1–7?)

- First, let's count the number of students in blue zones, star zones, and dotted zones.
- Next, let's count the number of students in blue star zones, blue dotted zones, and star dotted zones.
- What about the number of students in all three zones?

Class Activity (15 mins)



Question: How many students are there in zones **1–7**?

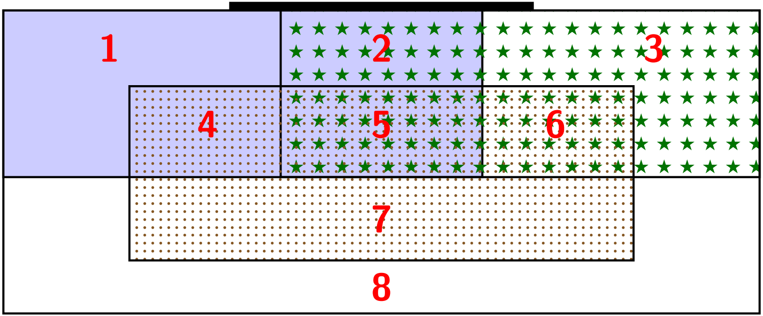
Answer: A first estimate..

$$\left(\begin{matrix}\text{Number in}\\ \textbf{blue}\text{ zones}\end{matrix}\right) + \left(\begin{matrix}\text{Number in}\\ \textbf{star}\text{ zones}\end{matrix}\right) + \left(\begin{matrix}\text{Number in}\\ \textbf{dotted}\text{ zones}\end{matrix}\right)$$

Some students **included** are counted more than once!

- Those in **blue star** zones are counted at least twice!
- Those in **blue dotted** zones are counted at least twice!
- Those in **star dotted** zones are counted at least twice!

Class Activity (15 mins)



Question: How many students are there in zones **1**–**7**?

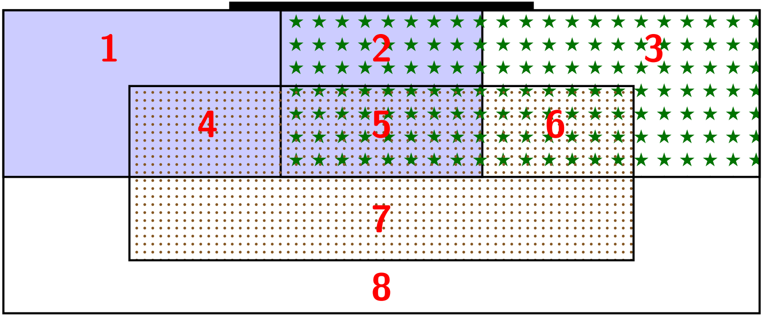
Answer: A better estimate..

$$\binom{\text{Number in}}{\textbf{blue}\text{ zones}} + \binom{\text{Number in}}{\textbf{star}\text{ zones}} + \binom{\text{Number in}}{\textbf{dotted}\text{ zones}}$$
$$- \binom{\text{Number in}}{\textbf{blue star}\text{ zones}} - \binom{\text{Number in}}{\textbf{blue dotted}\text{ zones}} - \binom{\text{Number in}}{\textbf{star dotted}\text{ zones}}$$

We have **excluded** all over-counted zones.

- Did we exclude/subtract too much?
- How many times is Zone **5** counted? **Answer:** 0 times!

Class Activity (15 mins)



Question: How many students are there in zones 1–7?

Answer: We have to **include** back those in **blue star dotted** zones:

$$\binom{\text{Number in}}{\textbf{blue}\text{ zones}} + \binom{\text{Number in}}{\textbf{star}\text{ zones}} + \binom{\text{Number in}}{\textbf{dotted}\text{ zones}}$$
$$- \binom{\text{Number in}}{\textbf{blue star}\text{ zones}} - \binom{\text{Number in}}{\textbf{blue dotted}\text{ zones}} - \binom{\text{Number in}}{\textbf{star dotted}\text{ zones}}$$
$$+ \binom{\text{Number in}}{\textbf{blue star dotted}\text{ zones}}$$

We found the exact value by a series of **inclusions** and **exclusions**.

Inclusion-exclusion principle

Inclusion-exclusion principle: Given any three sets $\mathbf{A}, \mathbf{B}, \mathbf{C}$,

$$\begin{aligned}|\mathbf{A} \cup \mathbf{B} \cup \mathbf{C}| = |\mathbf{A}| + |\mathbf{B}| + |\mathbf{C}| \quad & \text{(include all individual sets)} \\ - |\mathbf{A} \cap \mathbf{B}| - |\mathbf{A} \cap \mathbf{C}| - |\mathbf{B} \cap \mathbf{C}| \quad & \text{(exclude all pairs of sets)} \\ + |\mathbf{A} \cap \mathbf{B} \cap \mathbf{C}| \quad & \text{(include all triples of sets).}\end{aligned}$$

More generally, given any n sets $A_1, \ldots, A_n$,

$$\begin{aligned}|A_1 \cup \cdots \cup A_n| = {} & |A_1| + \cdots + |A_n| \\ & - |A_1 \cap A_2| - |A_1 \cap A_3| - \cdots - |A_{n-1} \cap A_n| \\ & + |A_1 \cap A_2 \cap A_3| + \cdots + |A_{n-2} \cap A_{n-1} \cap A_n| \\ & \ldots \\ & + (-1)^{n-1} |A_1 \cap A_2 \cap \cdots \cap A_n|.\end{aligned}$$

Note: The intersection of every k-tuple $(A_{i_1}, \ldots, A_{i_k})$ of sets is included if k is odd, and excluded if k is even.

Summary

- Counting methods
- Multiplication Rule and Tuples
- Permutations and Combinations
- Class Activity: Inclusion-exclusion principle

Reminder: Homework Set 1 is due next Cohort Class.