# MM 225: AI and Data Science

## Permutations, combinations and binomial

M P Gururajan and Hina A Gokhale

August 22, 2023

# Lecture 5: Permutations, combinations and binomial

# Outline

# The problem of points

Suppose G and H are playing several games. G needs three more wins and H needs two more wins to get the stakes. If the games are stopped at this point, what is the fair division of stakes? Assume that the stakes are Rs. 64.

# Pascal's and Fermat's solutions

- Pascal: Algorithmic
- Involved Pascal's identity
- Fermat: combinatoric
- This lecture: binomial (Pascal's identity) and permutations and combinations

# Birthday problem

# Questions

## Question

*In a class of 200 students, what is the probability that two students share the same birthday?*

## Question

*How big should a class be so that the probability of two students sharing the same birthday is more than 50%?*

# Guess

**Question**

*In a class of 200 students, what is the probability that two students share the same birthday?*

Can we guess that this will be about $200/365 \approx 0.548$

# Answer

Probablity is $\approx \frac{200}{365} = 0.55$

- **Link:** https://www.menti.com/alyv6higa18m

- **Code:** Go to menti.com and use code 5529 0593

# Results

▸ Results of the poll

# Sample space

- A sequence of 200 birthdays – assuming no repetition
- In other words, we find the probability that no two people have the same birthday
- We get the probability that at least two share a birthday by subtracting this number from unity!

# Number of sequences

- Assume nobody is born in a leap year
- There are 365 possible birthdays
- First person could be born on any of the 365 days
- Second person has only 364 possibilities
- Third person has only 363 possibilities
- ... and so on
- 200th person has $365 - 200 + 1 = 166$ possibilities
- What are all the number of sequences?

# Permutation

### Definition

Let A be any finite set. A permutation of A is a one-to-one mapping of A into itself.

### Theorem

*The total number of permutations of a set A of n elements is n!.*

### Definition

Let A be an n-element set and let k be an integer between 0 and n. Then a k-permutation of A is an ordered listing of a subset of A of size k.

### Theorem

*The total number of k-permutations of a set A of n elements is* $\frac{n!}{(n-k)!}$

# Calculation

- Assume nobody is born in a leap year
- There are 365 possible birthdays
- What is the probability that all the birth days are distinct?
- First person could be born on any of the 365 days
- Second person has 364 possibilities
- Third person has 363 possibilities ...
- 200th person has $365 - 200 + 1 = 166$ possibilities

# Probability

- Thus, for 200 students, the probability is $\frac{365 \cdot 364 \cdot 363 \cdots 166}{365^{200}}$
- The probability that two students have the same birthday is closer to 1
- Note: factorial and calculation of the same using Stirling's approximation

# Guess

**Question**

*How big should a class be so that the probability of two students sharing the same birthday is more than 50%?*

Can we guess this number and verify the guess using a python script?

# Binomial coefficients

# Combinations

- Consider a set $U$ with $n$ elements
- Let $\phi$ the null set, and the set $U$ be also considered as subsets of $U$
- What are the number of distinct subsets of $U$: $2^n$
- What are the number of distinct subsets of $U$ that has $j$ elements?
- Answered by combinations!

# Binomial coefficients

- Number of distinct subsets with $j$ elements that can be chosen with $n$ elements: $\binom{n}{j}$ ($n$ choose $j$)
- $\binom{n}{j}$: binomial coefficient
- Assume $n > 0$
- There is only one way to choose zero elements or $n$ elements
- For integers $j$ and $n$, with $0 < j < n$

**Theorem**

$\binom{n}{j} = \binom{n-1}{j} + \binom{n-1}{j-1}$

# Bernoulli trials

**Definition**

A Bernoulli trials process is a sequence of $n$ chance experiments such that each experiment has two possible outcomes with probabilities of $p$ for 'success' and $1-p$ for 'failure'

- What is the probability that in $n$ Bernoulli trials there are exactly $j$ successes?
- Probability $b(n, p, j) = \binom{n}{j} p^j q^{n-j}$

# Binomial distribution

### Definition

Let $n$ be a positive integer and $0 < p < 1$. The random variable B which counts the number of successes in a Bernoulli trials process with parameters $n$ and $p$ is distributed as binomial distribution $b(n, p, k)$

Use of binomial distribution: Tutorials. For example, to check the number of trials needed to determine the efficacy of a new medicine and simulation of Galton board

# A problem

G and H have given a quiz to 300 students in their data analysis course. The quiz consisted of 10 true or false questions. If the students tossed a coin to answer the questions, how many students would have scored ten on ten? How many zero on ten? If there are about 100 students who had scored 10 on 10 and about 5 scored 0 on 10, what can you conclude about the difficulty level of the quiz?

# Python to answer the question

- Probability $b(n,p,j) = \binom{n}{j} p^j q^{n-j}$
- $n = 10$; $p = 0.5$; $j = 0$ and $j = 10$
- ```
  from scipy.stats import binom
  binom.pmf(0,10,0.5)
  binom.pmf(10,10,0.5)
  ```
- Probability: $\approx 0.001$
  ```

# Python script

```
from scipy.stats import binom
import matplotlib.pyplot as plt

x = binom.rvs(10,0.5,size=300)

plt.hist(x,bins=[0,1,2,3,4,5,6,7,8,9,10])
plt.show()
```

# The icon problem

# Question

I very carefully arrange the app icons on my cell phone:

- Moive apps: Netflix, Google tv, Mubi, Prime Video, Disney-Hotstar, Zee 5
- Entertainment apps: Alexa, fire tv, amazon music, youtube, apple music, eppo music, pocket casts
- Call taxi /auto aggregators: Uber, Ola, Namma Yatri
- Food: Zwiggy, Zomato
- Frequent: camera, messages, whatsapp, gmail, maps, calendar, authenticator

My niece takes my phone and randomly arranges the icons. What is the probability that none of the icons end up at their original position? At least $k$ of them end up in their original position?

# Fixed points

- Permutations: one-to-one mapping of a set onto itself
- The points that map to themselves after random permutations: fixed points
- Library problem, hat check problem, ...
- How to solve these problems?
- Simulation! Certainly.
- Exact solution?

# Inclusion-Exclusion principle

## Theorem

*Let $P$ be a probability measure on a sample space $\Omega$ and let $(A_1, A_2, \ldots A_n)$ be a finite set of elements. Then,*

$$\begin{aligned} P(A_1 \cup A_2 \cup \cdots \cup A_n) &= \sum_{i=1}^{n} P(A_i) \\ &- \sum_{1 \le i < j \le n} P(A_i \cap A_j) \\ &+ \sum_{1 \le i < j < k \le n} P(A_i \cap A_j \cap A_k) \\ &- \cdots \end{aligned} \tag{1}$$

# Inclusion-Exclusion principle

- A generalisation of $P(A \cup B) = P(A) + P(B) - P(A \cap B)$
- To find the probability that at least one event out of $n$ events $A_i$ occurs, add the probability of each event, subtract the probabilities of all possible two-way interactions, add the probability of all three-way interactions and so on.

# Icon problem: solution using inclusion-exclusion principle

- $A$: ordered set of $n$ icons
- $A_i$: the event that the $i$-th element remains fixed under permutation – a one-to-one map of $A$ onto itself
- If $a_i$ is fixed, the remaining $(n-1)$ elements are arbitrarily mapped
- There are $(n-1)!$ such permutations
- $P(A_i) = \frac{(n-1)!}{n!} = \frac{1}{n}$
- There are n choices for $a_i$
- The first term in Eq. (1) is 1

$$\begin{aligned} P(A_1 \cup A_2 \cup \cdots \cup A_n) &= \sum_{i=1}^{n} P(A_i) \\ &- \sum_{1 \le i < j \le n} P(A_i \cap A_j) \\ &+ \sum_{1 \le i < j < k \le n} P(A_i \cap A_j \cap A_k) \\ &- \cdots \end{aligned}$$

# Other terms

- Let $(a_i, a_j)$ be fixed
- The permutations of remaining elements is $(n-1)!$
- Thus $P(A_i \cap A_j) = \frac{(n-2)!}{n!} = \frac{1}{n(n-1)}$
- There are $\binom{n}{2}$ such terms
- The second term in Eq. (1) is thus $\binom{n}{2}\frac{1}{n(n-1)} = \frac{n(n-1)}{2!}\frac{1}{n(n-1)} = \frac{1}{2!}$
- Similarly, the third term is $\frac{1}{3!}$

$$
\begin{aligned}
P(A_1 \cup A_2 \cup \cdots \cup A_n) \quad = \quad & \sum_{i=1}^{n} P(A_i) \\
& - \sum_{1 \le i < j \le n} P(A_i \cap A_j) \\
& + \sum_{1 \le i < j < k \le n} P(A_i \cap A_j \cap A_k) \\
& - \cdots
\end{aligned}
$$

# Solution

- P(at least one fixed point)

$$1 - \frac{1}{2!} + \frac{1}{3!} - \cdots (-1)^{n-1} \frac{1}{n!} \tag{2}$$

- P(no fixed point)

$$\frac{1}{2!} - \frac{1}{3!} + \cdots (-1)^{n} \frac{1}{n!} \tag{3}$$

- P(no fixed point) is the sum of the first $n$ terms of the expansion for $\frac{1}{e}$

# ‘No fixed point’ as a function of $n$

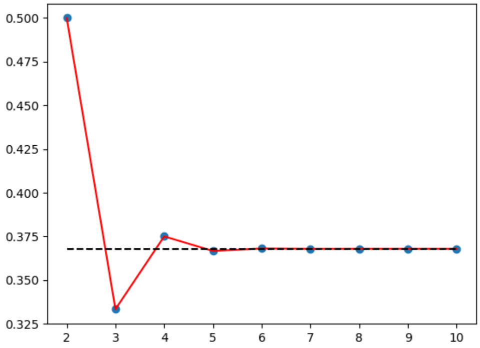


Figure: The dotted line corresponds to $\frac{1}{e}$.

# Summary

- Permutations
- Combinations
- Bernoulli trials and binomial distribution
- Next lecture: Conditional probability

THANK YOU!!!