Revision

Introduction to Statistics

- **Sample Space**
  - A sample space is a set that contains all outcomes of an experiment.

- **Event**
  - An **event is a subset of the sample space**. There is a technical restriction on what subsets can be events.
- **Disjoint events**
  - Two events with an empty intersection are said to be disjoint events.

Introduction to Probability

- Probability is a function P that assigns to each event a real number b/w 0 & 1.
- The entire probability space (sample space, events and probability function) should satisfy the following two axioms:
  - P(S) = 1 (Probability of entire sample space equals 1)
  - If $E_1, E_2, E_3, E_4 \ldots$ are disjoint events then,
    P($E_1 \cup E_2 \cup E_3 \cup E_4 \cup \ldots$) = P($E_1$) + P($E_2$) + P($E_3$) + P($E_4$) + $\ldots$.

Permutation and Combination

- Permutation and combination are *mathematical concepts* used to **count and arrange objects in different ways.**
- The *key difference* b/w them is whether the *order of arrangement* matters.

Permutations:

- Permutation refer to the **different ways in which a set of objects can be arranged in a specific order.**
- For a set of *n* distinct items taken *r* at a time , the number of permutations is given by

$$P(n,r) = \frac{n!}{(n-r)!}$$

  where '!' denoted factorial.
- e.g. - *Arranging 3 books (A,B,C)* on a shelf gives *6 permutations*: **ABC, ACB, BAC, BCA, CAB, CBA**.

Combinations:

- Combinations refer to the different ways in which a set of objects can be selected without considering the order.
- For a set of n distinct items taken r at a time, the number of combinations is given by

$$C(n, r) = \frac{n!}{r!(n-r)!}$$

- e.g. - Choosing 2 books out of 3 (A, B, C) gives 3 combinations: **AB, AC, BC**.

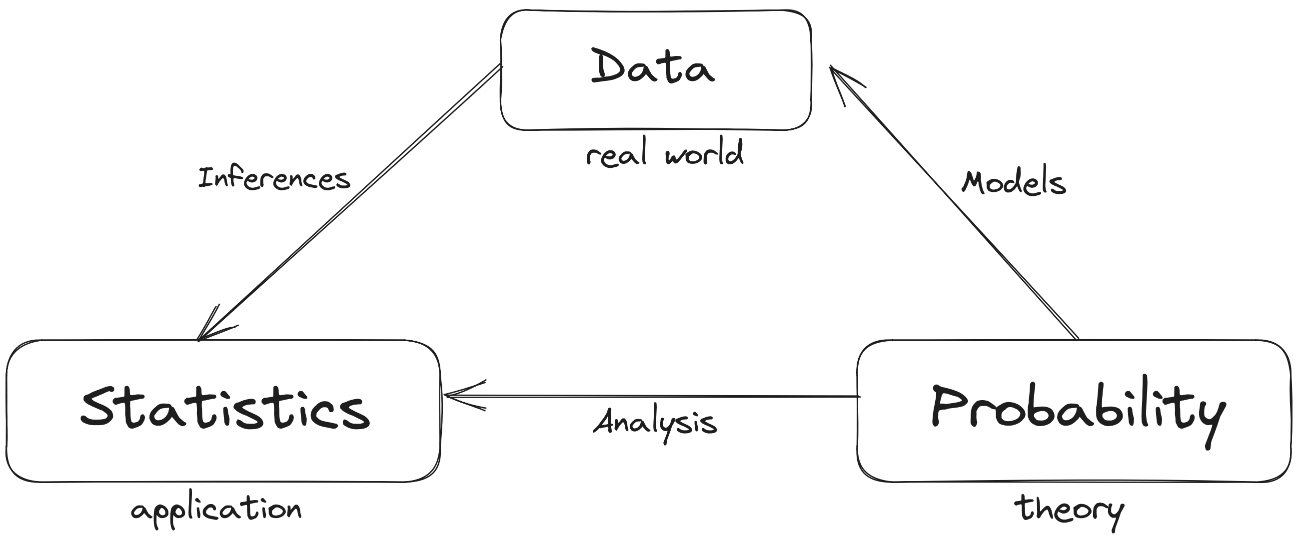
Summary

- Use *permutations* when the *order matters*, and you're concerned with different arrangements.
- Use *combinations* when the *order doesn't matter* and you're interested in selecting items without considering their order.

Week 0 Topics

Data, statistics and Probability

Statistical study of phenomena



Bernoulli Trials

- A Bernoulli trial is a **random experiment or event with only two possible outcomes**: *success or failure.*

- Each trial is **independent**, meaning that the outcome of one trail does not affect the outcome of any other trial.
- The *probability of success* (often denoted as **p**) *remains constant across all trials*; while the *probability of failure is **1-p**.*
- Examples of Bernoulli trials include flipping a coin, rolling a die for a specific outcome.

Single Bernoulli Trial

- A single Bernoulli trial refers to **one specific instance of a Bernoulli trial**.
- It involves a **single experiment or event** with two possible outcomes: **Success or Failure**.
- The **probability of success (p) and failure (1-p) remain constant** for that particular trial.
- e.g. - *Flip a coin with a fair, unbiased coin and the outcome of the coin toss represents a single Bernoulli trial*.
    - *The possible outcome of this single Bernoulli trial would be:*
        - *Success: The coin lands on head.*
        - *Failure: The coin lands on tails.*
    - In this case,
        - the probability of success (**p**) = $\frac{1}{2}$ = **0.5** &
        - probability of failure (**1-p**) = $1 - \frac{1}{2}$ = $\frac{1}{2}$ =**0.5**.

Repeated Bernoulli Trials

- Repeated Bernoulli trials involve **conducting multiple Bernoulli trials in succession**.
- **Each trial is independent** and the **probability of success (p) and failure(1-p) remains the same** for each trial.
- Sample space will consist of 2^n outcomes, where **n** is the no. of trials.
- e.g. - Toss a fair coin 3 times. Find the *probability of which no. of tails in all trials is 0* ?
    - Here n = 3
    - P(0 tails) = P(trial 1 is H and trial 2 is H and trial 3 is H) = $\frac{1}{2} * \frac{1}{2} * \frac{1}{2}$ = $\frac{1}{8}$

Tip

If a Bernoulli (p) trial is repeated n times independently, then

$$P(b_1, b_2, b_3 ... b_n) = p^w(1-p)^{n-w}$$

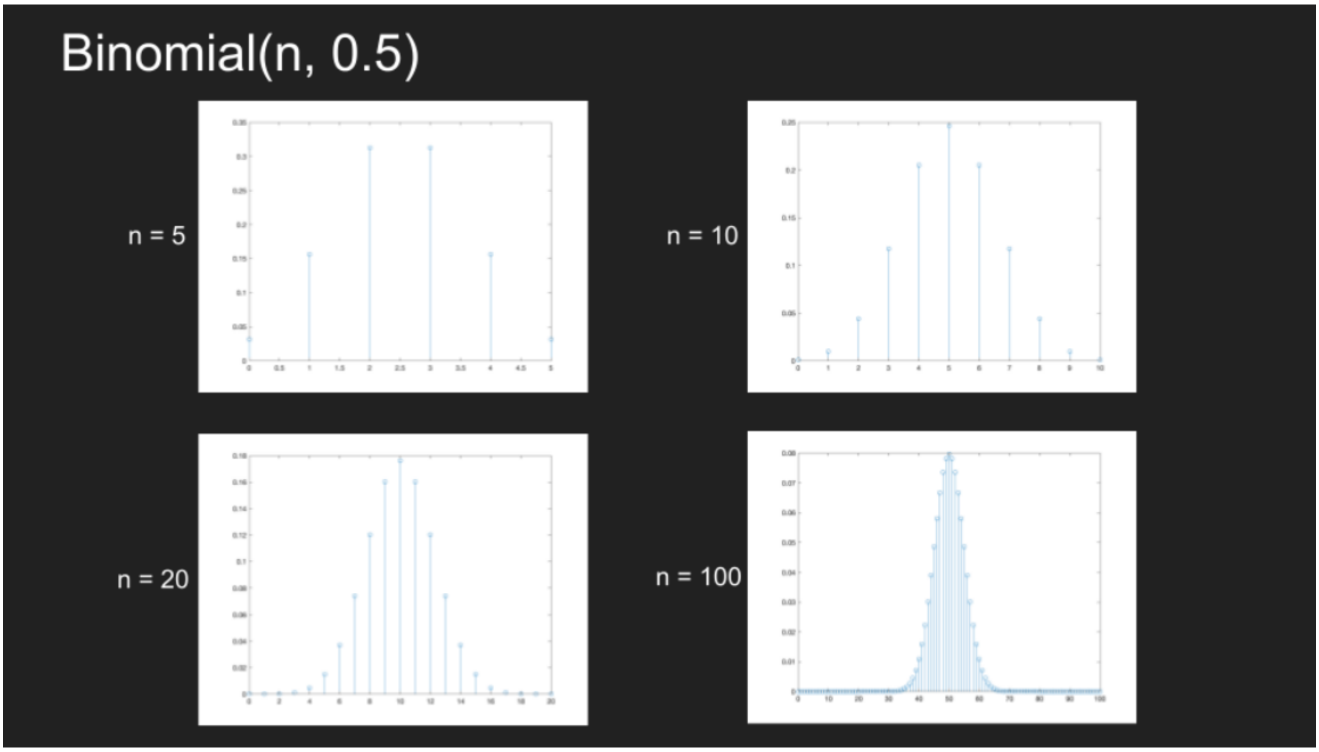
where w = Number of success in $b_1, b_2, b_3 ... b_n$

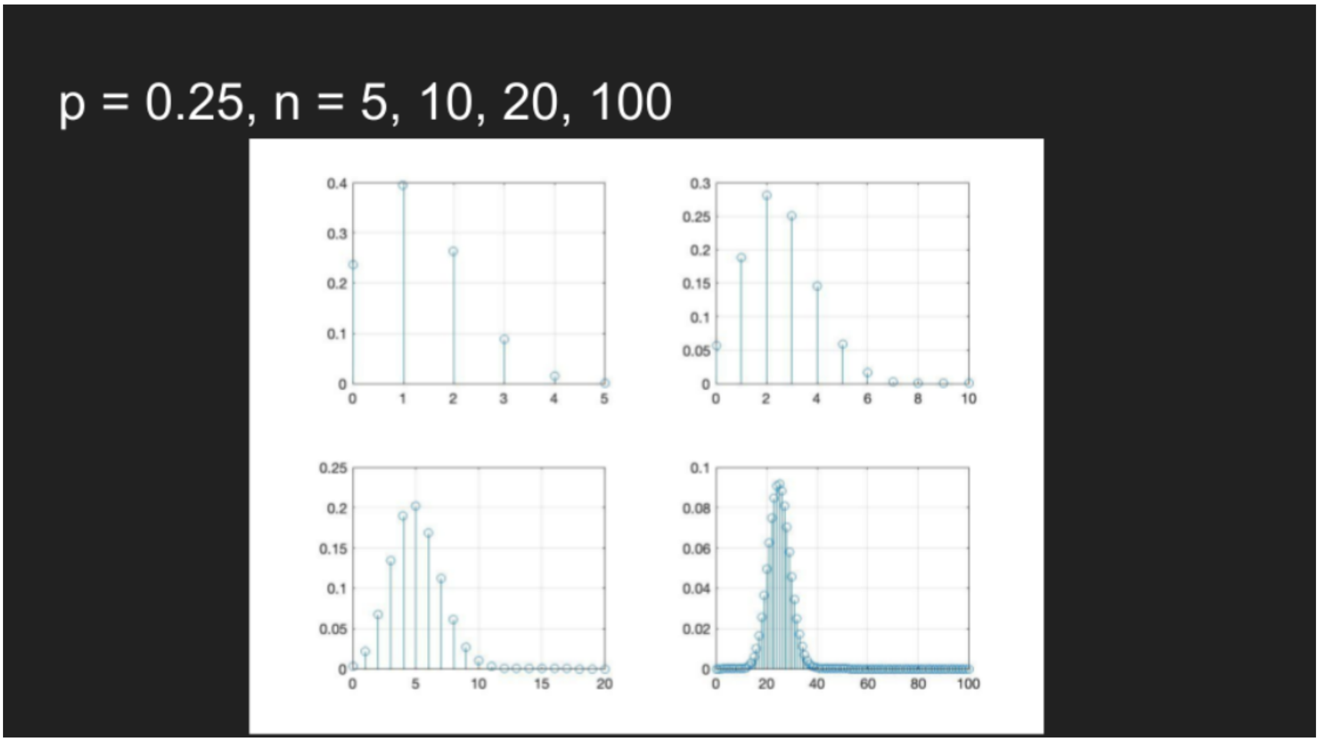
Binomial Distribution

- The binomial distribution is a *probability distribution* that describes *the number of successes in a fixed number of independent Bernoulli trials,* where each trial has only two possible outcomes: *success or failure*.
- **Probability Distribution function** is given by

$$P(B(n,p)=k) = {}^{n}C_k \cdot p^k(1-p)^{n-k} \quad ; \qquad k = 0,1,2,3,\ldots,n$$

Visualization of Binomial Distribution





Note

- The plot **starts at** $(1-p)^n$ and then **increases till** it **reaches the peak** and the **falls to** p^n**.**
- The **peak is roughly around *np*** and the exact values are as follows:
  - If ***'(n+1)p'* is an integer**, then it is **bimodal** (i.e. two peaks) and the **two peak values** are ***'(n+1)p'* and *'(n+1)p - 1'*.**
  - If ***'(n+1)p'* is not an integer**, then there exists **a unique modal value** (i.e. unique peak value) and it's the **integral part of *'(n+1)p'***.

Solved Examples

1. Each person has a disease with probability 0.1 independently. Out of 100 random persons tested for the disease, what is the probability that 20 persons test positive? Assume that the disease can be tested accurately with no false positives.
   - Solution:
     - Consider each test as a Bernoulli trial. Where
       - Probability of success = 0.1
       - Probability of failure = 1- 0.1 = 0.9
       - Here this Bernoulli trial is repeated 100 times independently. Hence to find the probability of getting 20 success in these 100 trial is given by:

$$P(B(n,p)=k) = {}^{n}C_k \cdot p^k(1-p)^{n-k} \quad ; \qquad k = 0,1,2,3,\ldots,n$$

       - n = 100; k = 20; p = 0.1; 1-p = 0.9

$$\begin{aligned} P(B(100,0.1)=20) &= {}^{100}C_{20} \cdot (0.1)^{20} \cdot (1-0.1)^{100-20} \quad ; \\ &= {}^{100}C_{20} \cdot (0.1)^{20} \cdot (0.9)^{80} \quad ; \\ &\approx \ 0.0012; \end{aligned}$$

Geometric Distribution

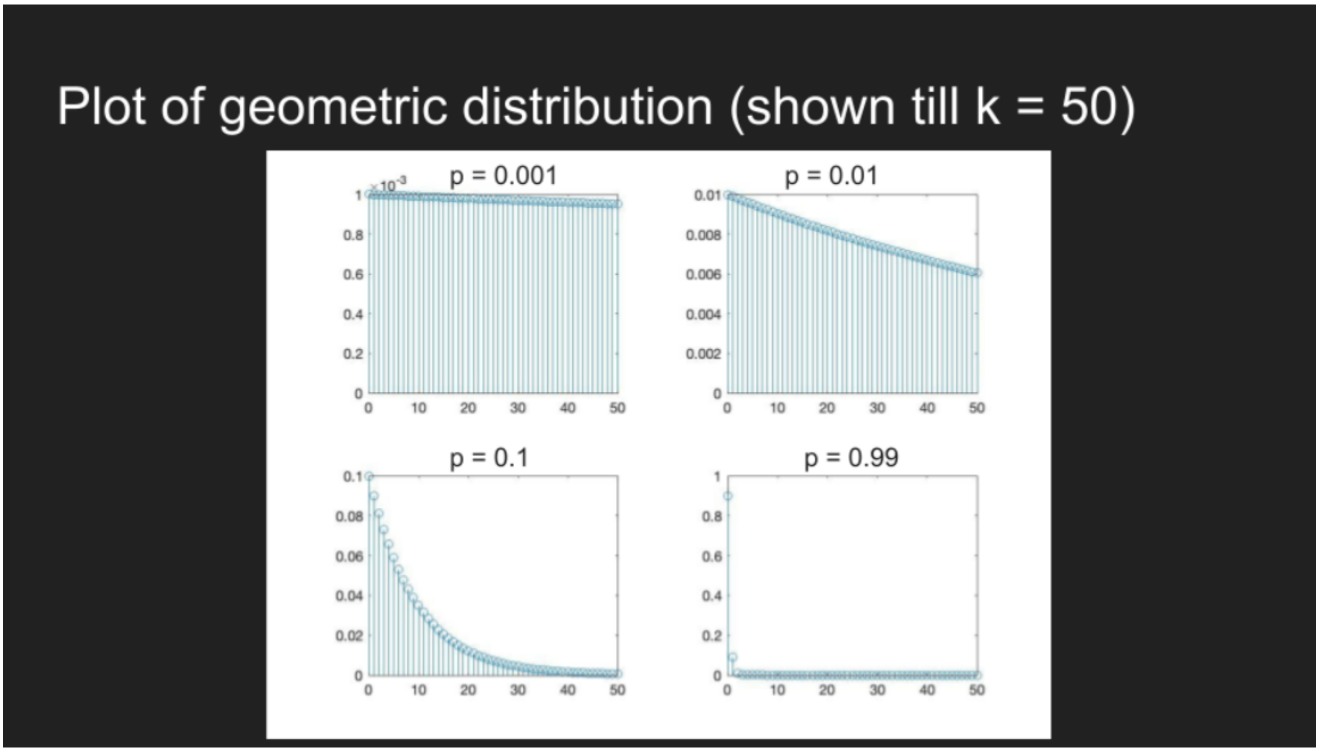
- **Geometric distribution** represents *the number of trials needed until the first success in a sequence of binary events (success or failure)*, where each trial has the same probability of success(p) and the trials are independent.
- Consider a **Bernoulli(*p*)** trial with *S = {0,1}*. If this **Bernoulli(*p*)** trial is repeated independently, then the number of trials needed for the first success is given by the Geometric distribution as follows:
  - Sample space; *S* = {1,2,3,4,6,....}, This is because we can get the first success in 1st trial, 2nd trial, 3rd trial..., and so on.

$$P(G(p)=k) = (1-p)^{k-1} \cdot p \quad ; \qquad k = 1,2,3,4,\ldots$$

- The probability of getting the first success **within the first k trials** is given by:

$$P(G(p) \leq k) = 1 - (1 - p)^k \quad ;$$

Visualization of Geometric Distribution



> ✎ **Note**
>
> - The plot **starts at** p and then it keeps falling.
> - Even though the plot keeps on decreasing but, if $p < 1$, it never goes all the way to zero.