Hypergeometric Distribution - Sai Shanmukh Sangubhotla

The **Hypergeometric Distribution** is a **discrete probability distribution** that gives the probability of obtaining exactly **k** successes in a sample of size **n** drawn **without replacement** from a finite population of size **N** containing exactly **K** successes.

Key points:

- The population is **finite**.
- Sampling is done **without replacement**.
- Each selected item is classified as either a **success** or a **failure**.
- It calculates the probability of getting **exactly k successes**.

example: A bag has 10 balls — 4 red, 6 blue. You draw 3 balls without putting any back. What's the probability you get exactly 2 red balls?

This is exactly the hypergeometric setup:

- $N = 10$ (total balls)
- $K = 4$ (red balls = "successes")
- $n = 3$ (draws)
- $k = 2$ (successes wanted)

Hypergeometric vs. Binomial Distribution

Binomial Distribution

The Binomial distribution models the number of successes in a **fixed number of independent trials**, where each trial has exactly two outcomes (success/failure) and the **probability of success stays the same every single time**.

This happens when you're sampling **with replacement**, or when trials are naturally independent of each other — like flipping a coin or rolling a die repeatedly. What happens on one trial has zero effect on the next.

Its formula is:P(X=k)=(kn)pk(1−p)n−k

where n is the number of trials and p is the constant probability of success. The mean is np and the variance is np(1−p)np(1-p) np(1−p).

Hypergeometric Distribution

The Hypergeometric distribution models the number of successes when you draw a **fixed-size sample from a finite population, without replacement**. Because you don't put items back, **the probability of success changes with every draw** — pulling out a success item means there's one fewer left for the remaining draws. The trials are **not independent**; each draw depends on what happened before it.

This is why the population size (N) actually matters here, unlike in Binomial. You need to know exactly how many total items and how many successes exist in that finite population.

Its formula is:P(X=k)=(nN)(kK)(n−kN−K)

where N is the population size, K is the number of successes in the population, and n is the sample size drawn. The mean is nNK and the variance is nNK⋅NN−K⋅N−1N−n — note the extra term at the end, called the finite population correction factor, which accounts for the fact that the population is limited and shrinking as you draw.

The Essential Distinction

Binomial assumes **replacement (or independence)** — the probability of success never changes. Hypergeometric assumes **no replacement** — the probability of success shifts after every draw because the population itself is shrinking and changing composition.

A simple test to tell them apart: ask whether removing an item changes the odds for the next draw. If yes, it's Hypergeometric. If no, it's Binomial.

When the population N is very large compared to the sample size n, removing a few items barely changes the proportions left behind — so Hypergeometric behaves almost exactly like Binomial in that case, with p = K/N.

A box contains 10 items, of which 4 are defective. If 3 items are drawn at random without replacement, find the probability that exactly 2 are defective.

Solution:
Here N = 10, K = 4, n = 3, k = 2.

P(X=2)=(42)(61)(103)=6×6120=36120=0.3P(X=2) =
P(X=2)=(310)(24)(16)=1206×6=12036=0.3

Answer: The probability of exactly 2 defective items is **0.3**.