PROB

This passage compares the **variance** of the number of red balls selected when sampling with and without replacement from a finite population.

1. **Sampling with Replacement:**
   - If sampling were done with replacement, the count of red balls would follow a **binomial distribution** with parameters n (sample size) and $p = \frac{A}{A+B}$ (proportion of red balls).
   - The **mean** of this count would be $n \cdot p = \frac{nA}{A+B}$, and the **variance** would be $n \cdot p \cdot (1 - p)$.
2. **Sampling without Replacement:**
   - When sampling without replacement, the count of red balls has a **different variance** due to the dependence between draws.
   - The variance formula for sampling without replacement is:
     $$\mathrm{Var}(X) = n \cdot p \cdot (1 - p) \cdot \frac{T - n}{T - 1}$$
     where $T = A + B$ is the total number of balls.
3. **Finite Population Correction (FPC):**
   - The term $\alpha = \frac{T-n}{T-1}$ is called the **finite population correction** (FPC) factor, which reduces the variance when sampling without replacement. This factor accounts for the decreasing pool of available red balls as selections are made without replacement.
   - Key points about α:
     - When $n = 1$, $\alpha = 1$, so there's no distinction between sampling with and without replacement for a single draw.
     - When $n = T$, $\alpha = 0$, resulting in a variance of 0, as all balls are selected, so there's no randomness left.
     - As T becomes very large compared to n, α approaches 1, making the variance in sampling without replacement close to the binomial variance of sampling with replacement. This shows that for a large population relative to the sample, both methods yield similar results.

Great question! Let's break down why sampling with replacement leads to a **binomial distribution** and why sampling without replacement gives a **hypergeometric distribution**.

1. Sampling with Replacement (Binomial Distribution)

- **Independence of Events**: When you sample with replacement, each draw is independent. For example, if you're drawing balls from a box and you put each one back after drawing, the probability of drawing a red or blue ball remains the same each time.
- **Fixed Probability for Each Trial**: In this case, each draw is like a Bernoulli trial (an event with only two possible outcomes, like red or blue). Because you're sampling with replacement, the probability of drawing a red ball, p, is constant across all trials.
- **Binomial Conditions**: Since we have a fixed number of trials n, each trial is independent, and each trial has the same probability of success p (like drawing a red ball), this process fits the conditions for a **binomial distribution**.

The binomial distribution $\text{Binomial}(n, p)$ describes the probability of a given number of successes in n independent trials with success probability p.

So, **sampling with replacement** will **always** yield a binomial distribution if the outcome is defined as the count of "successes" (like drawing red balls).

2. Sampling without Replacement (Hypergeometric Distribution)

- **Dependence Between Trials**: When you sample without replacement, the probability of drawing a red or blue ball changes with each draw because the total number of balls decreases. For example, if you draw a red ball and don't replace it, there are now fewer red balls and fewer total balls.
- **Changing Probability**: The probability of drawing a red or blue ball is no longer fixed with each trial. Each draw affects the next, making them **dependent** events.
- **Hypergeometric Conditions**: In this scenario, you have a finite population with a certain number of successes and failures (e.g., red and blue balls),

and you are drawing a sample of a specific size without replacement. This setup aligns with the **hypergeometric distribution**.

The hypergeometric distribution describes the probability of obtaining a specific number of successes (like red balls) in a sample of size \(n \) drawn without replacement from a population with a known number of successes and failures.

Key Differences and Applications

- **Independence vs. Dependence**: The primary difference lies in whether or not each trial is independent. With replacement, each trial is independent, making it binomial. Without replacement, each trial is dependent on the previous ones, making it hypergeometric.
- **Finite Population Correction (FPC)**: This correction only applies when sampling without replacement from a finite population, which decreases the variance as the sample size approaches the population size.

Will Sampling with Replacement Always Be Binomial?

- Yes, as long as each sample is independent and the probability of success remains the same, **sampling with replacement will always yield a binomial distribution** if we're counting the number of successes over \(n \) trials.
- However, if we're not dealing with binary outcomes (like counting red or blue balls), and if there are more than two types of outcomes, we might need other distributions (e.g., **multinomial** distribution for more than two categories).

In summary:

- **Sampling with replacement** → **Binomial** (constant probability, independent trials).
- **Sampling without replacement** → **Hypergeometric** (changing probability, dependent trials).

Poisson

✅ The limitation of the Poisson distribution refers to its ability to predict an unlimited number of events, which can lead to unrealistic scenarios. In contrast, the binomial distribution has a fixed maximum number of successes (n), making it more suitable for situations where there's a clear limit on possible outcomes. So, while the Poisson distribution allows for very large counts, this flexibility can complicate its application in real-world cases where an upper limit exists.