STAT 2593

Lecture 014 - The Hypergeometric and Negative Binomial Distributions

Dylan Spicker

The Hypergeometric and Negative Binomial Distributions

Learning Objectives

1. Understand the hypergeometric distribution, its use cases, and its properties.

2. Understand the negative binomial distribution, its use cases, and its properties.


BETHLEHEM

The Hypergeometric Distribution

- If we are sampling from a finite population, **without** replacement, where each item is a success or failure, this is captured by the **hypergeometric distribution**.

The Hypergeometric Distribution

- If we are sampling from a finite population, **without** replacement, where each item is a success or failure, this is captured by the **hypergeometric distribution**.
  - We assume that there is a fixed number of trials, n.

The Hypergeometric Distribution

- If we are sampling from a finite population, **without** replacement, where each item is a success or failure, this is captured by the **hypergeometric distribution**.
  - We assume that there is a fixed number of trials, n.
  - We assume that there are total number of items, N.

The Hypergeometric Distribution

- If we are sampling from a finite population, **without** replacement, where each item is a success or failure, this is captured by the **hypergeometric distribution**.
  - We assume that there is a fixed number of trials, n.
  - We assume that there are total number of items, N.
  - We assume that a fixed number of these items are successes, M.

The Hypergeometric Distribution

- If we are sampling from a finite population, **without** replacement, where each item is a success or failure, this is captured by the **hypergeometric distribution**.
  - We assume that there is a fixed number of trials, n.
  - We assume that there are total number of items, N.
  - We assume that a fixed number of these items are successes, M.
- The random variable X of the number of successes then follows $X \sim \text{HyperGeo}(N, M, n)$.

The Hypergeometric Distribution

- If we are sampling from a finite population, **without** replacement, where each item is a success or failure, this is captured by the **hypergeometric distribution**.
  - We assume that there is a fixed number of trials, n.
  - We assume that there are total number of items, N.
  - We assume that a fixed number of these items are successes, M.
- The random variable X of the number of successes then follows $X \sim \text{HyperGeo}(N, M, n)$.
  - We have $E[X] = n\frac{M}{N}$, $\text{var}(X) = \frac{N-n}{N-1} n \frac{M}{N}\left(1 - \frac{M}{N}\right)$, and

$$p(x) = \frac{\binom{M}{x}\binom{N-M}{n-x}}{\binom{N}{n}}.$$

The Hypergeometric and Binomial Distributions

- Note that if we take $p = \frac{M}{N}$ then $E[X] = np$ and $\mathrm{var}(X) = \frac{N-n}{N-1} np(1-p)$.

The Hypergeometric and Binomial Distributions

- Note that if we take $p = \frac{M}{N}$ then $E[X] = np$ and $\text{var}(X) = \frac{N-n}{N-1} np(1-p)$.

- This is almost identical to the Binomial distribution, with an extra multiplicative term.

The Hypergeometric and Binomial Distributions

- Note that if we take $p = \frac{M}{N}$ then $E[X] = np$ and $\text{var}(X) = \frac{N-n}{N-1} np(1-p)$.
- This is almost identical to the Binomial distribution, with an extra multiplicative term.
  - This term is known as the **finite population correction factor**.

The Hypergeometric and Binomial Distributions

- Note that if we take $p = \frac{M}{N}$ then $E[X] = np$ and $\text{var}(X) = \frac{N-n}{N-1} np(1-p)$.

- This is almost identical to the Binomial distribution, with an extra multiplicative term.

  - This term is known as the **finite population correction factor**.

  - As $N \to \infty$, then this tends towards 1, which demonstrates why the binomial can approximate sampling without replacement.

The Negative Binomial Distribution

- If you have repeated Bernoulli trials, and you count the number of trials until a set number of successes, the corresponding random variable follows a **negative binomial distribution**.

The Negative Binomial Distribution

- If you have repeated Bernoulli trials, and you count the number of trials until a set number of successes, the corresponding random variable follows a **negative binomial distribution**.
  - We assume that there is a fixed number of success to observe, r.

The Negative Binomial Distribution

- If you have repeated Bernoulli trials, and you count the number of trials until a set number of successes, the corresponding random variable follows a **negative binomial distribution**.
  - We assume that there is a fixed number of success to observe, r.
  - We assume that there is a constant probability of success, p.

The Negative Binomial Distribution

- If you have repeated Bernoulli trials, and you count the number of trials until a set number of successes, the corresponding random variable follows a **negative binomial distribution**.
  - We assume that there is a fixed number of success to observe, r.
  - We assume that there is a constant probability of success, p.
  - We assume that each trial is independent of each other.

The Negative Binomial Distribution

- If you have repeated Bernoulli trials, and you count the number of trials until a set number of successes, the corresponding random variable follows a **negative binomial distribution**.
  - We assume that there is a fixed number of success to observe, r.
  - We assume that there is a constant probability of success, p.
  - We assume that each trial is independent of each other.
- The random variable X of the counts of the number of failures before the rth success follows $X \sim \text{NegBin}(r, p)$.

The Negative Binomial Distribution

- If you have repeated Bernoulli trials, and you count the number of trials until a set number of successes, the corresponding random variable follows a **negative binomial distribution**.
  - We assume that there is a fixed number of success to observe, r.
  - We assume that there is a constant probability of success, p.
  - We assume that each trial is independent of each other.
- The random variable X of the counts of the number of failures before the rth success follows $X \sim \text{NegBin}(r, p)$.
  - We have $E[X] = \frac{r(1-p)}{p}$, $\text{var}(X) = \frac{r(1-p)}{p^2}$, and

$$p(x) = \binom{x+r-1}{r-1} p^r (1-p)^x.$$

The Negative Binomial and Geometric Distributions

- If we take $r = 1$ then the negative binomial is just the geometric distribution.

The Negative Binomial and Geometric Distributions

- If we take $r = 1$ then the negative binomial is just the geometric distribution.
    - Very often we care about $r = 1$, which is why it gets a special name and is treated separately.

The Negative Binomial and Geometric Distributions

- If we take $r = 1$ then the negative binomial is just the geometric distribution.
  - Very often we care about $r = 1$, which is why it gets a special name and is treated separately.
- Like the geometric distribution, sometimes the random variable X will instead count the total number of trials, including successes.

Summary

- The hypergeometric distribution characterizes the number of successes when sampling *without* replacement.
- The hypergeometric distribution relies on the assumptions of fixed, finite total, with a fixed number of successes, and a fixed number of trials.
- The negative binomial distribution counts the number of failures required to achieve a certain number of successes.
- The negative binomial distribution relies on the same assumptions regarding bernoulli trials as we have seen.
- Both distributions have closed form PMFs, expectations, and variances.