Discrete Random Variables

Chapter 3: Discrete Random Variables

Anduo Wang
Temple University
Email: anduo.wang@gmail.com
https://cis-linux1.temple.edu/∼tug29203/25fall-2033/index.html

Hypergeometric Distribution

- You have a bag that contains b blue marbles and r red marbles.
- Choose $k \leq b + r$ marbles at random without replacement.
- Let X be the number of blue marbles in your sample.
- By definition, $X \leq min(k, b)$ and number of red marbles in your sample must be less than or equal to r, so we have $X \geq max(0, k - r)$.
- $R_X = \{max(0, k-r), max(0, k-r)+1, max(0, k-r)+2, ..., min(k, b)\}$.
- Total number of ways to choose k marbles from $b + r$ marbles is $\binom{b+r}{k}$.

Hypergeometric Distribution

- The number of ways to choose x blue marbles and $k-x$ red marbles is $\binom{b}{x}\binom{r}{k-x}$.
- In general we have

A random variable X is said to be a *Hypergeometric* random variable with parameters b, r and k, shown as $X \sim Hypergeometric(b, r, k)$, if its range is R_X, and its PMF is given by

$$P_X(x) = \begin{cases} \frac{\binom{b}{x}\binom{r}{k-x}}{\binom{b+r}{k}} & \text{for } x \in R_X \\ 0 & \text{otherwise} \end{cases}$$

Expectation

- The expectation for different special distributions
  - $X \sim \text{Bernoulli}(p)$: $EX = p$.
  - $X \sim \text{Geometric}(p)$: $EX = \frac{1}{p}$.
  - $X \sim \text{Poisson}(\lambda)$: $EX = \lambda$.
  - $X \sim \text{Binomial}(n, p)$: $EX = np$.
  - $X \sim \text{Pascal}(m, p)$: $EX = \frac{m}{p}$.
  - $X \sim \text{Hypergeometric}(b, r, k)$: $EX = \frac{kb}{b+r}$.

Derive the expected value of $X \sim Hypergeometric(b, r, k)$

- The total number of blue items included in k draws is the number of blue items obtained on the first draw plus the number of blue items obtained on the second draw plus ... plus the number of special items obtained on the k-th draw.

$$X = I_1 + I_2 + \cdots + I_k$$

- I_j is the indicator function: I_j on the $j-$th draw is 1 (resp. 0) if the $j-$th draw produces a blue (resp. red) item.

$$E(I_1) = P(I_1 = 1) = \frac{b}{b+r}$$

$$E(I_1) = E(I_2) = \cdots = E(I_k) = \frac{b}{b+r} \; (why?)$$

Ordered Sampling without Replacement, Revisited

- If a random ordered sample of size k is drawn from a population of size N, then on any particular one of the k draws each of the N items has the same probability $\frac{1}{N}$ of appearing.
  - Hint: $\frac{P_{k-1}^{N-1}}{P_k^N}$

Binomial Approximation to Hypergeometric

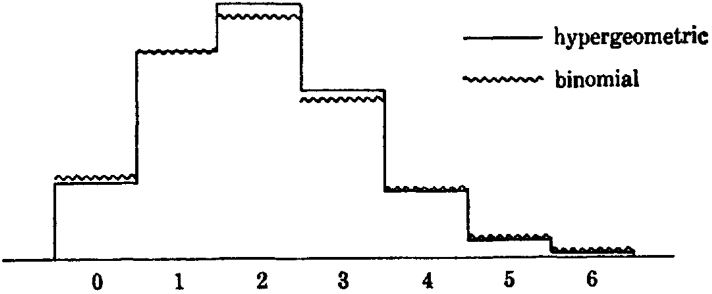
- An approximation for the hypergeometric distribution (Hypergeometric(b, r, k)) when the sample constitutes only a small fraction of the population
  - $k/b + r$ is small

Binomial Approximation to Hypergeometric

- An approximation for the hypergeometric distribution (Hypergeometric(b, r, k)) when the sample constitutes only a small fraction of the population
  - $k/b + r$ is small
- Suppose that a sample of k items is drawn **with replacement** from the population of $b + r$ items
  - the chance of getting a blue item is $\frac{b}{b+r}$, just as it was for sampling without replacement, but now the k draws are unrelated
  - Form a sequence of s binomial trials, and the number of blue items in the sample has the binomial distribution with $n = k$ and $p = \frac{b}{b+r}$.

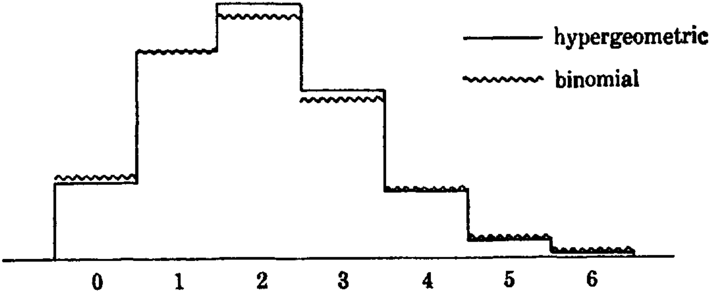
Binomial approximation to hypergeometric

$X \sim hypergeometric(k = 10, b = 20, r = 80)$ and
$Y \sim Bin(n = 10, p = .2)$



Binomial approximation to hypergeometric

$X \sim hypergeometric(k = 10, b = 20, r = 80)$ and
$Y \sim Bin(n = 10, p = .2)$



- Common practife to use the binomial distribution for the computation of hypergeometric probabilities