# Expectation

Let $X$ follow a hypergeometric distribution in which $n$ objects are selected from $N$ objects with $m$ of the objects being one type, and $N - m$ of the objects being a second type. What is the expected value of $X$?

# Expectation

Let $X$ follow a hypergeometric distribution in which $n$ objects are selected from $N$ objects with $m$ of the objects being one type, and $N - m$ of the objects being a second type. What is the expected value of $X$?

$$E(X) = \frac{nm}{N}$$

# Expectation- proof

$$E(X) = \sum_x x \frac{\binom{m}{x}\binom{N-m}{n-x}}{\binom{N}{n}}$$

Now, $\binom{m}{x} = \frac{m!}{x!(m-x)!}$, and, $\binom{N}{n} = \frac{N!}{n!(N-n)!} = \frac{N(N-1)!}{n \cdot (n-1)!(N-n)!} = \frac{N}{n} \cdot \frac{(N-1)!}{(n-1)!(N-1-(n-1))!} = \frac{N}{n} \cdot \binom{N-1}{n-1}$

Hence, $E(X) = \sum_x x \frac{\frac{m \cdot (m-1)!}{x!(m-x)!}\binom{N-m}{n-x}}{\frac{N}{n} \cdot \binom{N-1}{n-1}} =$

$$\sum_x \frac{nm}{N} \frac{\frac{(m-1)!}{(x-1)!(m-1-(x-1))!}\binom{(N-1)-(m-1)}{(n-1)-(x-1)}}{\binom{N-1}{n-1}} = \frac{nm}{N}$$

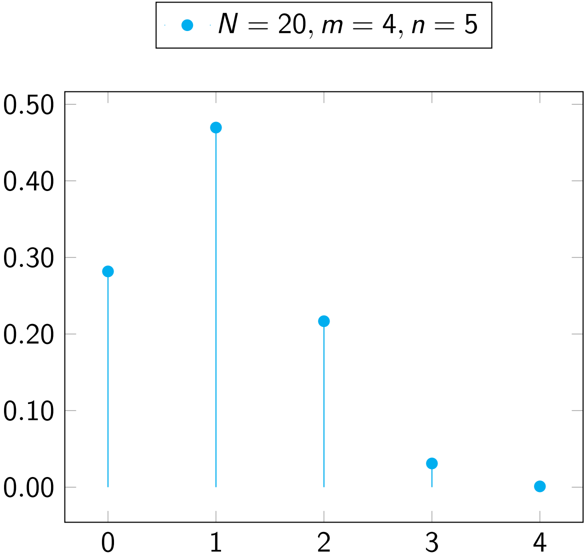
# Variance

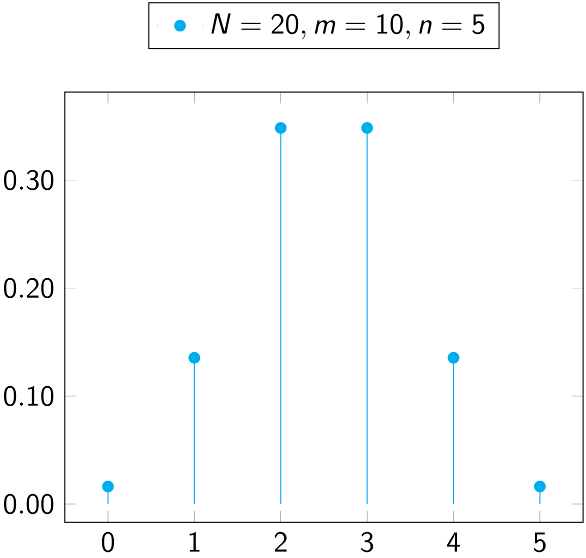
Let $X$ follow a hypergeometric distribution in which $n$ objects are selected from $N$ objects with $m$ of the objects being one type, and $N - m$ of the objects being a second type. What is the variance of $X$?

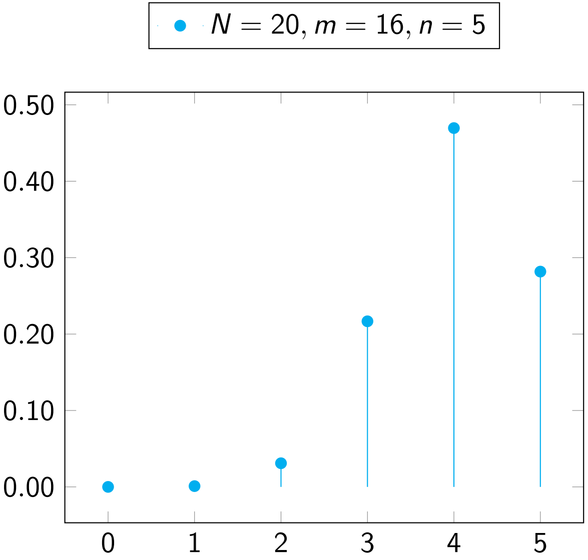
# Variance

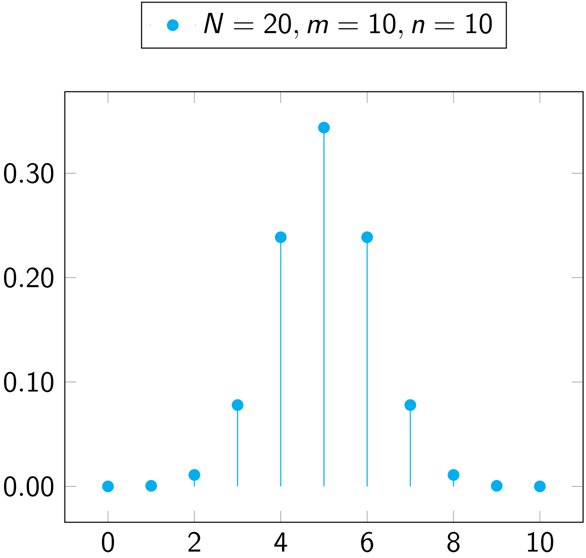
Let $X$ follow a hypergeometric distribution in which $n$ objects are selected from $N$ objects with $m$ of the objects being one type, and $N - m$ of the objects being a second type. What is the variance of $X$?

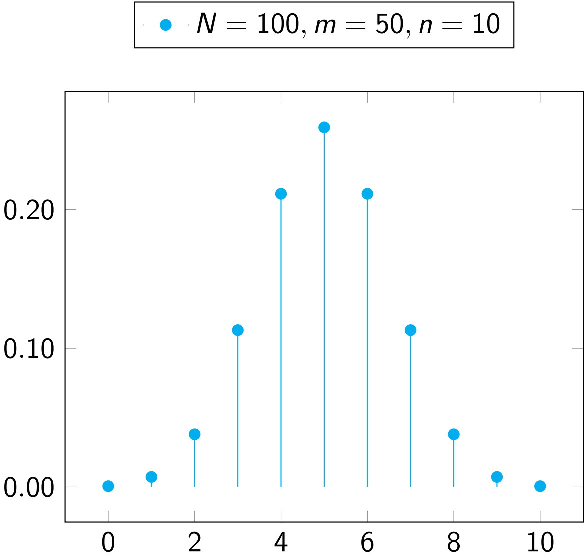
$$Var(X) = n\frac{m}{N}\frac{N-m}{N}\frac{N-n}{N-1}$$

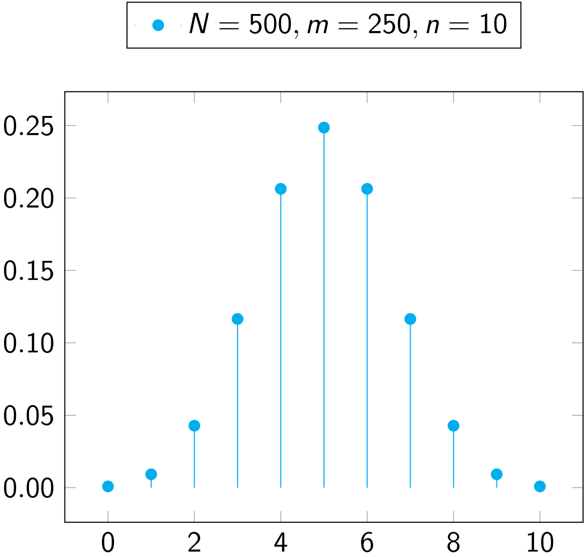

$N = 20, m = 4, n = 5$
0.50
0.40
0.30
0.20
0.10
0.00
0
1
2
3
4


$N = 20, m = 10, n = 5$
0.30
0.20
0.10
0.00
0
1
2
3
4
5


N = 20, m = 16, n = 5
0.50
0.40
0.30
0.20
0.10
0.00
0
1
2
3
4
5


$N = 20, m = 10, n = 10$
0.30
0.20
0.10
0.00
0
2
4
6
8
10

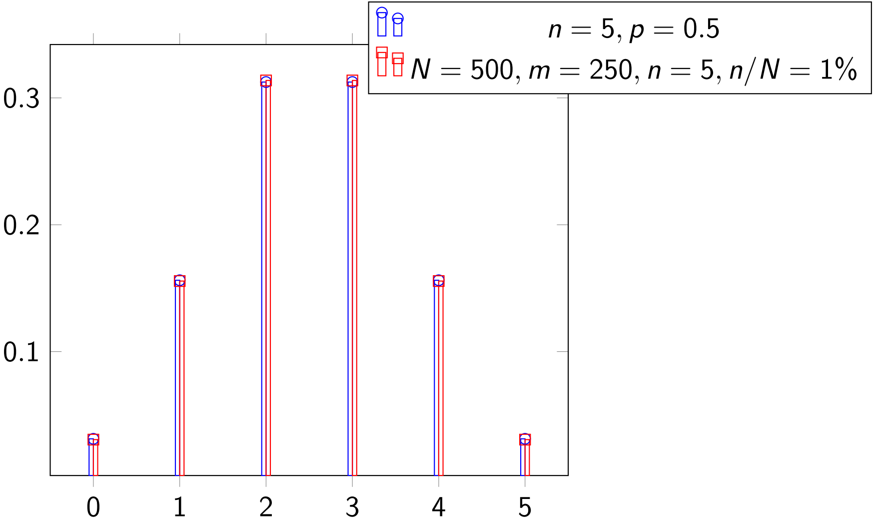

$N = 100, m = 50, n = 10$
0.20
0.10
0.00
0
2
4
6
8
10


$N = 500, m = 250, n = 10$
0.25
0.20
0.15
0.10
0.05
0.00
0
2
4
6
8
10

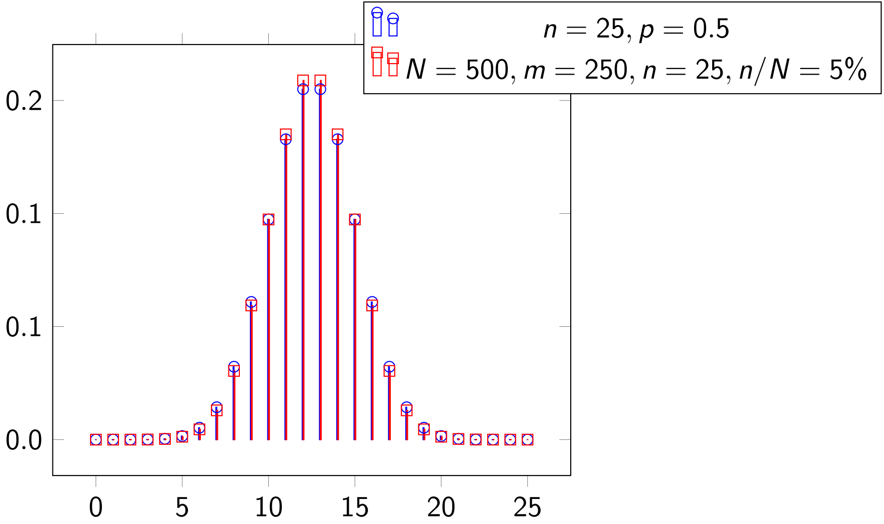
# Expectation and variance

- $X \sim Hypergeometric(N, m, n)$
  - $E(X) = \dfrac{nm}{N}$
  - $Var(X) = n\frac{m}{N}\frac{N-m}{N}\frac{N-n}{N-1}$
- $Y \sim Bin\left(n, \dfrac{m}{N}\right)$
  - $E(X) = \dfrac{nm}{N}$
  - $Var(X) = n\frac{m}{N}\frac{N-m}{N}$
- $\frac{N-n}{N-1}$ is known as finite population correction
  - For $n = 1$, replacement has no effect both are Bernoulli trial
  - For $n = N$, the whole population is sampled- hence variance is zero.
- If the population $N$ is very large compared to the sample size $n$ (i.e. $N >> n$) then Hypergeometric$(N, m, n)$ is about Binomial$(n, \dfrac{m}{N})$.

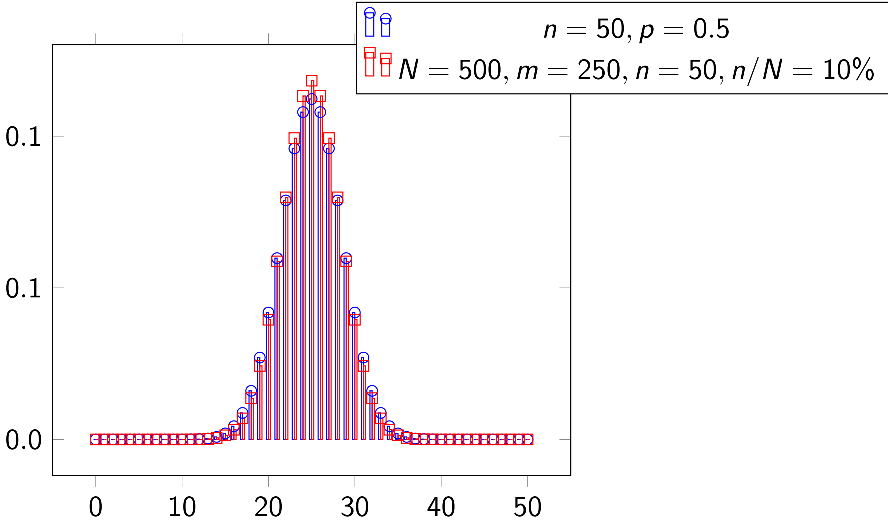
# Binomial versus Hyergeometric distribution



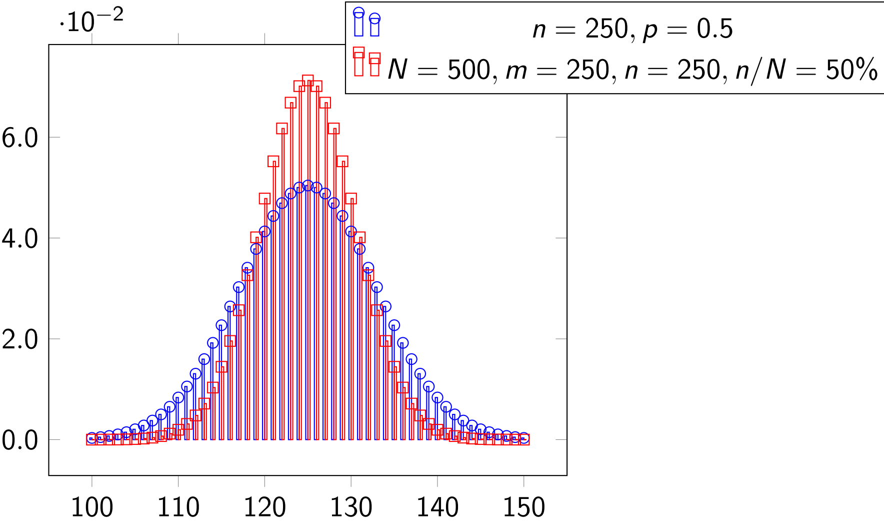
# Binomial versus Hyergeometric distribution



# Binomial versus Hyergeometric distribution



# Binomial versus Hypergeometric distribution



# Section summary

- Binomial versus Hypergeometric