XV. Hypergeometric Distribution

Hypergeometric Distribution

- The hypergeometric distribution describes choosing a committee of n men and women from a larger group of r women and $N-r$ men.
- This is an unordered choice, without replacement.
- What are the chances of getting exactly y women on our committee?
- $Y :=$ number of women on our committee

Formula for the Hypergeometric Distribution

- Fixed parameters:

$$\begin{aligned} N &:= \text{total number of people} \\ r &:= \text{number of women} \\ N-r &= \text{number of men} \\ n &:= \text{number on our committee} \end{aligned}$$

- Probability distribution:

$$\boxed{p(y) = \frac{\binom{r}{y}\binom{N-r}{n-y}}{\binom{N}{n}}, 0 \leq y \leq \min\{r, n\}}$$

Easy to remember.
If $r < n$, then we can only get up to r women.

Key Properties of the Hypergeometric Distribution

- **Mean**:

$$\boxed{\mu = E(Y) = \frac{nr}{N}}$$

- **Variance**:

$$\boxed{\sigma^2 = V(Y) = \frac{nr}{N}\frac{N-r}{N}\frac{N-n}{N-1}}$$

- **Standard deviation**:

$$\boxed{\sigma = \sqrt{V(Y)} = \sqrt{\frac{nr}{N}\frac{N-r}{N}\frac{N-n}{N-1}}}$$

Careful: These are honest fractions, not binomial coefficients.

Example I

There are 33 students in a class, 12 women and 21 men. We pick a committee of 7 students at random. What is the chance that the committee will contain exactly 5 women?

$$\begin{aligned} p(y) &= \frac{\binom{r}{y}\binom{N-r}{n-y}}{\binom{N}{n}} \\ &= \boxed{\frac{\binom{12}{5}\binom{21}{2}}{\binom{33}{7}}} \end{aligned}$$

Example II

What is the expected number of women on the committee in Example I?

$$\begin{aligned}\mu = E(Y) &= \frac{nr}{N} \\ &= \frac{7 \cdot 12}{33} = \boxed{\frac{28}{11} \text{ women}}\end{aligned}$$

Example III

Your shoe closet contains 10 pairs of shoes. Packing for a move, you begin throwing shoes into a box at random. The box fills up at 13 shoes. What is the probability that there are 5 left shoes and 8 right shoes in the box?

$$\begin{aligned}N &:= \text{total number of shoes} = 20 \\ r &:= \text{number of left shoes} = 10 \\ N - r &= \text{number of right shoes} = 10 \\ n &:= \text{number in the box} = 13 \\ p(y) &= \frac{\binom{r}{y}\binom{N-r}{n-y}}{\binom{N}{n}} \\ p(5) &= \boxed{\frac{\binom{10}{5}\binom{10}{8}}{\binom{20}{13}}}\end{aligned}$$

Example IV

What is the expected number of left shoes in the box in Example III?

$$\begin{aligned} \mu = E(Y) &= \frac{nr}{N} \\ &= \frac{13 \cdot 10}{20} = \boxed{6.5 \text{ left shoes}} \end{aligned}$$

Example V

Use indicator variables and linearity of expectation to prove that the expected value of a hypergeometric random variable is $\mu = \dfrac{nr}{N}$.

$$
\begin{aligned}
Y_1 &:= \text{\# of women picked in the first choice} = \begin{cases} 0 & \text{if we pick a man} \\ 1 & \text{if we pick a woman} \end{cases} \\
Y_2 &:= \text{\# of women picked in the second} \\
&\vdots \\
Y_n &:= \text{\# of women picked in the } n\text{th} \\
Y &:= \text{total number of women } = Y_1 + \cdots + Y_n
\end{aligned}
$$

(Recall that expectation is linear <u>even if the variables aren't independent</u>, which they surely aren't here: If you get a woman the first time, that is, if $Y_1 = 1$, then it is less likely that $Y_2 = 1$.)

$$
\begin{aligned}
E(Y_1) &:= \sum_{y=0,1} y p(y) \\
&= 0 \cdot P(\text{man}) + 1 \cdot P(\text{woman}) = \frac{r}{N} \\
E(Y) &= E(Y_1) + \cdots + E(Y_n) \\
&= \frac{r}{N} + \cdots + \frac{r}{N} \\
&= \boxed{\frac{nr}{N}}
\end{aligned}
$$