Math 170E

Lecture Notes Section 2.5 [*†]

Hypergeometric distribution

Instructor: Swee Hong Chan

NOTE: Materials that appear in the textbook but do not appear in the lecture notes might still be tested. Please send me an email if you find typos.

---

[*]Version date: Friday 22nd January, 2021, 20:15.

[†]This notes is based on March Boedihardjo's and Jamie Haddock's notes from the past quarters, and I would like to thank them for their generosity. "*Nanos gigantum humeris insidentes* (I am but a dwarf standing on the shoulders of giants)".

1 Hypergeometric distribution: Example

In a class of 40 students, 30 students are right-handed and 10 students are left-handed.

Your friendly instructor selects 4 students uniformly at random .

Let X be the number of right-handed students in these 4 students.

Then

$$
\begin{aligned}
P(X=0) &= P(\text{all 4 students are left-handed}) \\
&= \frac{\binom{30}{0}\binom{10}{4}}{\binom{40}{4}}. \\
P(X=1) &= P(\text{1 right-handed, 3 left-handed}) \\
&= \frac{\binom{30}{1}\binom{10}{3}}{\binom{40}{4}} \\
P(X=2) &= P(\text{2 right-handed, 2 left-handed}) \\
&= \frac{\binom{30}{2}\binom{10}{2}}{\binom{40}{4}}. \\
P(X=3) &= P(\text{3 right-handed, 1 left-handed}) \\
&= \frac{\binom{30}{3}\binom{10}{1}}{\binom{40}{4}}. \\
P(X=4) &= P(\text{4 right-handed, 0 left-handed}) \\
&= \frac{\binom{30}{4}\binom{10}{0}}{\binom{40}{4}}.
\end{aligned}
$$

2 Hypergeometric distribution: Definition

- There are N_1 objects of type A, and N_2 objects of type B;
- We choose n objects randomly out of these $N_1 + N_2$ objects;
- **Hypergeometric random variable** X is the number of chosen objects that are of type A.

3 Hypergeometric distribution: Mathematical definition

Let N_1, N_2, and n be fixed positive number.

The **hypergeometric random variable** X with **parameters** N_1, N_2, n has support and pmf

$$\begin{aligned} S &= \{0, 1, 2, \ldots, n\}; \\ f(x) &= \frac{\binom{N_1}{x}\binom{N_2}{n-x}}{\binom{N_1+N_2}{n}} \qquad \text{for } x \in \{0, 1, 2, \ldots, n\}. \end{aligned}$$

4 Hypergeometric distribution: Mean and variance

Theorem 1. *The hypergeometric random variable with parameters N_1, N_2, and n has mean and variance*

$$\begin{aligned} \mu &= n\,\frac{N_1}{N_1+N_2}; \\ \sigma^2 &= n\,\frac{N_1 N_2 (N_1+N_2-n)}{(N_1+N_2)^2 (N_1+N_2-1)}. \end{aligned}$$