Beginning of March 17, 2021

0.1 Hypergeometric Random Variables

Let N be the size of a population, m a number of "distinguished elements", and n the sample size. For example, let $N = 100, m = 50$, and $n = 10$. Then the **hypergeometric random variable** X gives the probability that, among a sample of n elements, exactly $X = x$ elements are "distinguished" and the remaining not "distinguished".

We write $X \sim H(N, m, n)$ where N, m, n are the parameters.

The range of X is given by $R(X) = \{\max(m - (N - n), 0), \dots, \min(m, n)\}$, but in this class we only consider the scenario where $m + n \leqslant N$ and so $R(X) = \{0, \dots, \min(m, n)\}$. Typically this would simply be $\{0, \dots, n\}$.

It follows naturally from combinatorics that

$$P_X(x) = \frac{\binom{m}{x}\binom{N-m}{n-x}}{\binom{N}{n}}.$$

The mean and variance are given by

$$E[X] = n \cdot \frac{m}{N} \text{ and } \sigma_X^2 = n \cdot \frac{m}{N}\left(1 - \frac{m}{N}\right)\left(\frac{N-n}{N-1}\right).$$

We will show how to derive these later (once we cover indicator random variable). Now simply notice the similaries of $E[X]$ and $\mathrm{Var}[X]$ of hypergeometric random variables to binomial. The $(N - n)/(N - 1)$ is called the **small population correction**. As $N \to \infty$ and n fixed, this correction $\to 1$ and indeed this looks more like a binomial distribution. (Note the only difference is that hypergeometric random variables are without replacements but binomials are with replacement.)

0.2 Discrete Uniform Random Variable

This is a probability distribution with a finite number of values that are equally likely to be observed. If there are k total values then each has a probability $1/k$. In general we write $X \sim \mathrm{unif}(a, b, n)$ but here we first consider the simplest case $X \sim \mathrm{unif}(0, 1, n)$. Then $R(X) = \{1/n, 2/n, \dots, (n - 1)/n\}$ (assuming $n \geqslant 2$). It immediately follows that $P_X(x) = 1/(n - 1)$.

Mean, Variance, and MGF

It's intuitive that the mean is 1/2:

$$\begin{aligned} E[X] &= \sum_x xP(x) = \sum_{i=1}^{n-1} \frac{i}{n}\frac{1}{n-1} \\ &= \frac{1}{n(n-1)}\sum_{i=1}^{n-1} i = \frac{1}{n(n-1)}\frac{n(n-1)}{2} = \frac{1}{2}. \end{aligned}$$

The second central moment is given by

$$\begin{aligned} E[X^2] &= \sum_x x^2P(x) = \sum_{i=1}^{n-1} \frac{i^2}{n^2}\frac{1}{n-1} \\ &= \frac{1}{n^2(n-1)}\sum_{i=1}^{n-1} i^2 = \frac{2n-1}{6n}. \end{aligned}$$

Therefore, the variance of X is given by

$$\mathrm{Var}[X] = \sigma_X^2 = E[X^2] - E[X]^2 = \frac{n-2}{12n}.$$