# The Hypergeometric Distribution

## *Math 394*

We detail a few features of the Hypergeometric distribution that are discussed in the book by Ross

## 1 Moments

Let $P[X = k] = \dfrac{\left(\begin{array}{c} m \\ k \end{array}\right)\left(\begin{array}{c} N-m \\ n-k \end{array}\right)}{\left(\begin{array}{c} N \\ n \end{array}\right)}$ (with the convention that $\left(\begin{array}{c} l \\ j \end{array}\right) = 0$ if $j < 0$, or $j > l$.

We detail the recursive argument from Ross. Consider, for $k = 1, 2, \ldots$

$$E[X^r] = \sum_{k=0}^{n} k^r \frac{\left(\begin{array}{c} m \\ k \end{array}\right)\left(\begin{array}{c} N-m \\ n-k \end{array}\right)}{\left(\begin{array}{c} N \\ n \end{array}\right)} \tag{1}$$

The sum can also be extended from 1 to $n$, since the term with $k = 0$ does not contribute to the result. Observe that

$$k\left(\begin{array}{c} m \\ k \end{array}\right) = k\frac{m!}{k!(m-k)!} = k\frac{m\cdot(m-1)!}{k\cdot(k-1)!(m-k)!} = m\frac{(m-1)!}{(k-1)!(m-1-[k-1])!} = m\left(\begin{array}{c} m-1 \\ k-1 \end{array}\right)$$

and also, as a consequence,

$$\left(\begin{array}{c} N \\ n \end{array}\right) = \frac{1}{n}\cdot n\cdot\left(\begin{array}{c} N \\ n \end{array}\right) = \frac{1}{n}N\left(\begin{array}{c} N-1 \\ n-1 \end{array}\right)$$

Inserting both identities in (1), after writing $k^r = k^{r-1}\cdot k$,

$$E[X^r] = \frac{nm}{N}\sum_{k=1}^{n} k^{r-1}\frac{\left(\begin{array}{c} m-1 \\ k-1 \end{array}\right)\left(\begin{array}{c} N-m \\ n-k \end{array}\right)}{\left(\begin{array}{c} N-1 \\ n-1 \end{array}\right)}$$

We can now change the summation index from $k-1,\ldots,n$ to $k-1 = j = 0,\ldots,n-1$, and notice that $n-k = n-(j+1) = (n-1)-j$, while $N-m = (N-1)-(m-1)$, so that

$$E[X^r] = \frac{nm}{N}\sum_{j=0}^{n-1}(j+1)^{r-1}\frac{\left(\begin{array}{c} m-1 \\ j \end{array}\right)\left(\begin{array}{c} (N-1)-(m-1) \\ (n-1)-j \end{array}\right)}{\left(\begin{array}{c} N-1 \\ n-1 \end{array}\right)}$$

Now, the red part of the formula can be compared with (1) to show that calling $Y$ a hypergeometrically distributed random variable with parameters $N-1, m-1, n-1$,

$$E\left[X^k\right] = \frac{nm}{N}E\left[(Y+1)^{r-1}\right]$$

This allows a simple recursion:

- $E[X] = \frac{nm}{N}$
- $E\left[X^2\right] = \frac{nm}{N} E[Y+1] = \frac{nm}{N}\left(\frac{(n-1)(m-1)}{N-1} + 1\right)$

Consequently,

$$Var[X] = E\left[X^2\right] - (E[X])^2 = \frac{nm}{N}\left(\frac{(n-1)(m-1)}{N-1} + 1\right) - \left(\frac{nm}{N}\right)^2 =$$
$$= \frac{nm}{N}\left(\frac{(n-1)(m-1)}{N-1} + 1 - \frac{nm}{N}\right)$$

Notice that, if we set $\frac{m}{N} = p$, and let $N, m \to \infty$, with $\frac{m}{N}$ fixed (that is, if we assume $n \ll N, m$) this expression tends to $np(1 = p)$, the variance of a binomial $(n, p)$. Incidentally, even without taking the limit, the expected value of a hypergeometric random variable is also $np$.

## 2 The Binomial Distribution as a Limit of Hypergeometric Distributions

The connection between hypergeometric and binomial distributions is to the level of the distribution itself, not only their moments. Indeed, consider hypergeometric distributions with parameters $N, m, n$, and $N, m \to \infty, \frac{m}{N} = p$ fixed. A random variable with such a distribution is such that

$$P[X = k] = \frac{\binom{m}{k}\binom{N-m}{n-k}}{\binom{N}{n}} = \frac{m!}{(m-k)!k!} \cdot \frac{(N-m)!}{(n-k)!(N-m-n+k)!} \cdot \frac{n!(N-n)!}{N!} =$$
$$= \binom{n}{k} \frac{m!}{(m-k)!} \cdot \frac{(N-m)!}{(N-m-n+k)!} \cdot \frac{(N-n)!}{N!} =$$
$$\binom{n}{k} \frac{m(m-1)(m-2)\cdots(m-k+1)}{N(N-1)(N-2)\cdots(N-k+1)} \cdot \frac{(N-m)(N-m-1)\cdots(N-m-(n-k)+1)}{(N-k)(N-k-1)\cdots(N-n+1)}$$

The color coding should help in keeping track where the various terms are coming from. In the last denominator, note also that $N - n + 1 = N - k - (n - k) + 1$. Let's take the limit now. We notice that:

$$\frac{m(m-1)(m-2)\cdots(m-k+1)}{N(N-1)(N-2)\cdots(N-k+1)} \approx \frac{m^k}{N^k} = p^k$$

since $k$ is kept finite while $m$ and $N$ diverge. Also, by the same argument, that is $n$ is kept finite,

$$\frac{(N-m)(N-m-1)\cdots(N-m-(n-k)+1)}{(N-k)(N-k-1)\cdots(N-n+1)} \approx \frac{(N-m)^{n-k}}{N^{n-k}} = \left(\frac{N-m}{N}\right)^{n-k} = (1-p)^{n-k}$$

Adding it all up, we have found that

$$\lim_{N,m\to\infty \, \frac{m}{N}=p} P[X = k] = \binom{n}{k} p^k (1-p)^{n-k}$$