



Math for the people, by the people.

## proof of variance of the hypergeometric distribution

Canonical name	ProofOfVarianceOfTheHypergeometricDistribution
Date of creation	2013-03-22 13:27:41
Last modified on	2013-03-22 13:27:41
Owner	mathwizard (128)
Last modified by	mathwizard (128)
Numerical id	13
Author	mathwizard (128)
Entry type	Proof
Classification	msc 62E15

We will first prove a useful property of binomial coefficients. We know

$$\binom{n}{k} = \frac{n!}{k!(n-k)!}.$$

This can be transformed to

$$\binom{n}{k} = \frac{n}{k} \frac{(n-1)!}{(k-1)!(n-1-(k-1))!} = \frac{n}{k} \binom{n-1}{k-1}. \quad (1)$$

The variance  $\text{Var}[X]$  of  $X$  is given by:

$$\text{Var}[X] = \sum_{x=0}^n \left( x - \frac{nK}{M} \right)^2 \frac{\binom{K}{x} \binom{M-K}{n-x}}{\binom{M}{n}}.$$

We expand the right hand side:

$$\begin{aligned} \text{Var}[X] &= \sum_{x=0}^n \frac{x^2 \binom{K}{x} \binom{M-K}{n-x}}{\binom{M}{n}} \\ &\quad - \frac{2nK}{M} \sum_{x=0}^n \frac{x \binom{K}{x} \binom{M-K}{n-x}}{\binom{M}{n}} \\ &\quad + \frac{n^2 K^2}{M^2} \sum_{x=0}^n \frac{\binom{K}{x} \binom{M-K}{n-x}}{\binom{M}{n}}. \end{aligned}$$

The second of these sums is the expected value of the hypergeometric distribution, the third sum is 1 as it sums up all probabilities in the distribution. So we have:

$$\text{Var}[X] = -\frac{n^2 K^2}{M^2} + \sum_{x=0}^n \frac{x^2 \binom{K}{x} \binom{M-K}{n-x}}{\binom{M}{n}}.$$

In the last sum for  $x = 0$  we add nothing so we can write:

$$\text{Var}[X] = -\frac{n^2 K^2}{M^2} + \sum_{x=1}^n \frac{x^2 \binom{K}{x} \binom{M-K}{n-x}}{\binom{M}{n}}.$$

Applying equation (??) and  $x = (x-1) + 1$  we get:

$$\text{Var}[X] = -\frac{n^2 K^2}{M^2} + \frac{nK}{M} \sum_{x=1}^n \frac{(x-1) \binom{K-1}{x-1} \binom{M-K}{n-x}}{\binom{M-1}{n-1}} + \frac{nK}{M} \sum_{x=1}^n \frac{\binom{K-1}{x-1} \binom{M-K}{n-x}}{\binom{M-1}{n-1}}.$$

Setting  $l := x - 1$  the first sum is the expected value of a hypergeometric distribution and is therefore given as  $\frac{(n-1)(K-1)}{M-1}$ . The second sum is the sum over all the probabilities of a hypergeometric distribution and is therefore equal to 1. So we get:

$$\begin{aligned}
\text{Var}[X] &= -\frac{n^2 K^2}{M^2} + \frac{nK(n-1)(K-1)}{M(M-1)} + \frac{nK}{M} \\
&= \frac{-n^2 K^2(M-1) + Mn(n-1)K(K-1) + KnM(M-1)}{M^2(M-1)} \\
&= \frac{nK(M^2 + (-K-n)M + nK)}{M^2(M-1)} \\
&= \frac{nK(M-K)(M-n)}{M^2(M-1)} \\
&= n \frac{K}{M} \left(1 - \frac{K}{M}\right) \frac{M-n}{M-1}.
\end{aligned}$$

This is the one we wanted to prove.