Hypergeometric

s = Hypergeometric(SubPop,Trials,Pop) = Hypergeometric(D,n,M)

Hypergeometric equations

Crystal Ball parameter restrictions

The Hypergeometric(D/M, *n, M*) distribution models the number of items of a particular type that there are in a sample of size *n* where that sample is drawn from a population of size *M* of which *D* are also of that particular type. Examples of the Hypergeometric distribution are shown below:
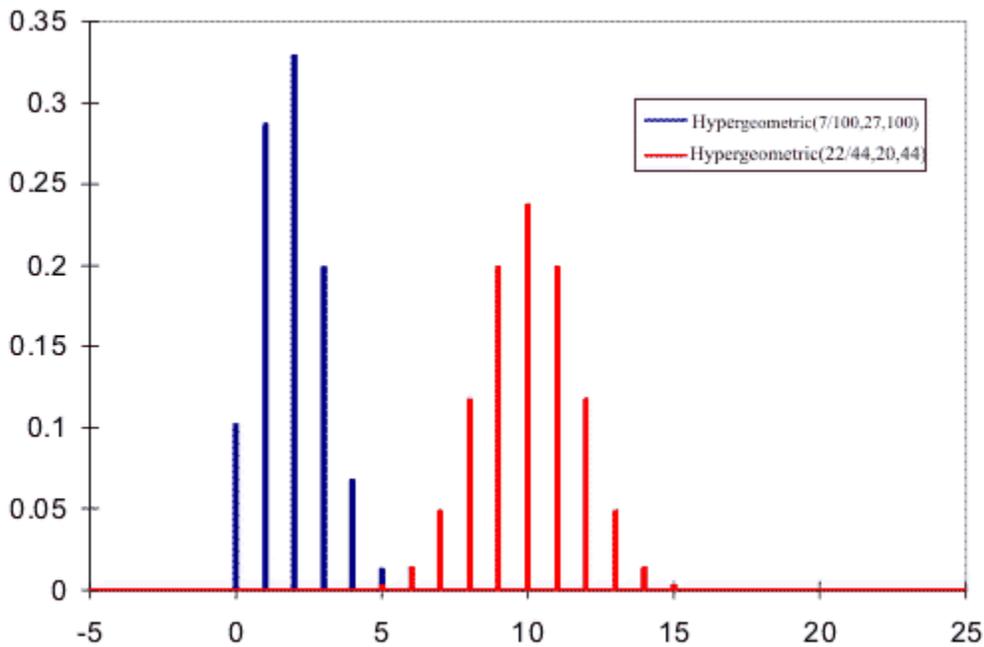
Examples

Tiles

A company has a stock of 2000 tiles which is known to contain 70 tiles that were not fired properly and will probably crack when exposed to the weather. The tiles are all mixed together and the inferior ones unfortunately cannot be visually identified. A customer orders 800 tiles. The number of faulty tiles he will receive can be estimated by Hypergeometric(70/2000, 800, 2000).

Capture-release-recapture experiment to estimate population size

An example of using the Hypergeometric distribution and Bayes' Theorem to estimate the number of tigers on an island is shown in the section uncertainty about a population size. Several animals are captured and tagged, then released back to the wild. Some time later, another set of animals is captured. The proportion that have tags provide a means, via Bayes' Theorem, to estimate the total population assuming complete diffusion of the tagged sample into the population.

Comments

The mathematics behind the Hypergeometric distribution assumes sampling from the population without replacement, which becomes more significant a restriction the closer the sample size *n* gets to the population size *M*. Where *n* is small compared to *M* and *D* (the general guideline is that $n < 0.1*M$), the Hypergeometric(D/M, *n, M*) looks very similar to Binomial(p,*n*) where $p = D/M$. The Hypergeometric distribution is closely related to the Inverse Hypergeometric distribution.

Although the first input for Crystal Ball's Hypergeometric Distribution is named "Prop", which stands for "probability", this is not a probability that is the same for every trial as is the case in the Binomial process, but it the probability for the first trial only. This is explained in more detail here.

The Excel function HYPGEOMDIST(x,n,D,M) returns the hypergeometric probability mass function.