Number of samples to get a specific s

Consider the situation where we are sampling without replacement from a population *M* with *D* items with the characteristic of interest until we have *s* items with the required characteristic. The distribution of the number of trials we will need to get *s* success can be easily calculated in the same manner as we developed the Negative Binomial distribution. The probability of observing (*s*-1) successes in (*x*-1) trials (i.e. we have had (x-1) - (s-1) = x -s failures) is given by direct application of the Hypergeometric distribution:

$$p(x, s-1) = \frac{\binom{D}{s-1}\binom{M-D}{x-s}}{\binom{M}{x-1}}$$

The probability *p* of then observing a success in the next trial (the *x*th trial), is simply the number of *D* items remaining (=*D*-(*s*-1)= D-s+1) divided by the size of the population remaining (= *M*-(*x*-1)=M-x+1):

$$p = \frac{(D-s+1)}{(M-x+1)}$$

and the probability of needing exactly *x* trials to obtain *s* success, where trials are stopped at the *s*th success, is then the product of these two probabilities:

$$p(x, s) = \frac{\binom{D}{s-1}\binom{M-D}{x-s}(D-s+1)}{\binom{M}{x-1}(M-x+1)}$$

This is the probability mass function for the Inverse Hypergeometric distribution InvHyperGeo(*s,D,M*) and is analogous to the Negative Binomial distribution for the binomial process and the Gamma distribution for the Poisson process. So:

n = InvHyperGeo(*s*,*D*,*M*)

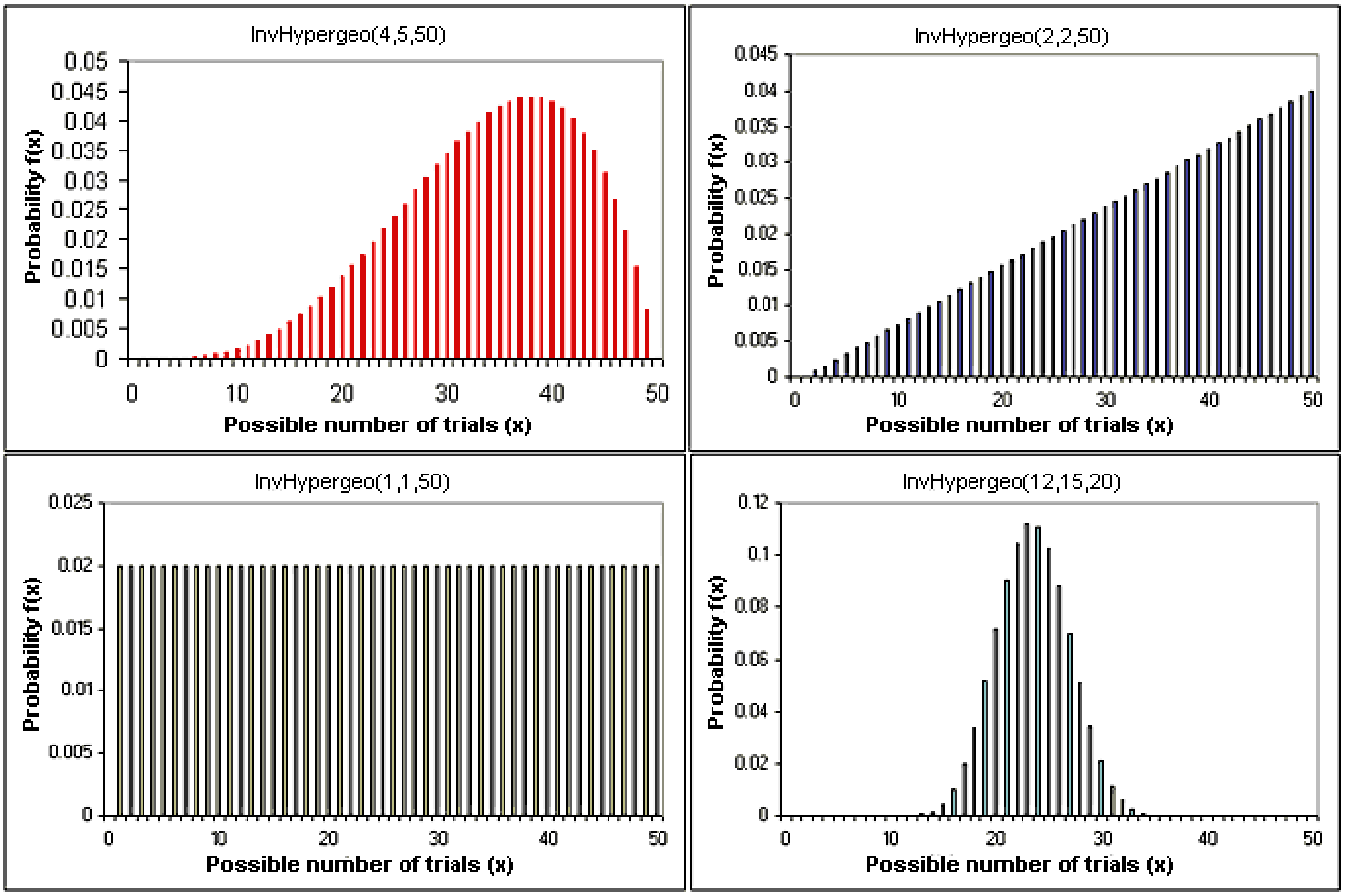
For a population *M* that is large compared to *s*, the Inverse Hypergeometric distribution is closely approximated by the Negative Binomial:

InvHypergeo(*s,D,M*) » NegBinomial(*D*/*M,s*)

and if the probability *D*/*M* is very small:

InvHypergeo(*s,D,M*) » Gamma(s,*M*/*D,s*)

The four figures below show examples of the Inverse Hypergeometric distribution. In the first figure you can see the probability mass function of the number of trials needed for getting 4 successes when drawing samples from a population 50 in which 5 individuals have the characteristic you are interested in. We leave to you the task to explain in words the figures 2 – 4.



An Inverse Hypergeometric distribution is sometimes called a *Negative Hypergeometric* distribution.