Hypergeometric Distribution (Isabella, Mari)

- *What is the question you came up with for this scenario?*
  What is the probability of selecting 50% transformed vs untransformed colonies on a plate with 400 total colonies if we pick 20 colonies?

- *More generally, what kind of problem is the function typically used for?*
  Probability of selecting x individuals belonging to group 1 without replacement given n total selections in a finite population of size N

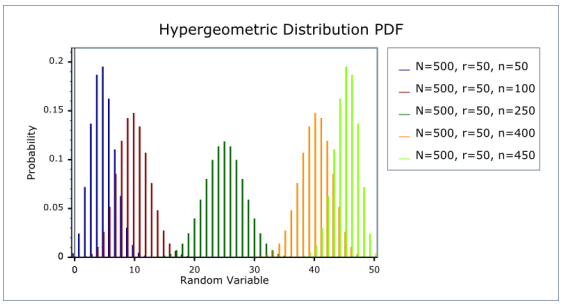
- *How is the distribution parameterized?*

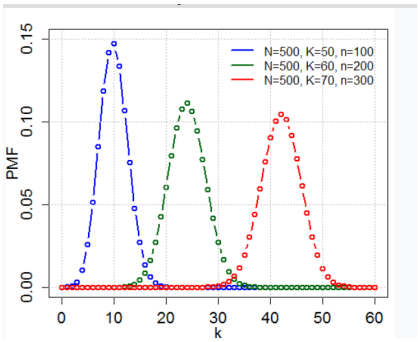
  Pdf formula - $\binom{M}{x}\binom{N-M}{n-x}\Big/\binom{N}{n}$
  (choose successes)*(choose failures) / (choose sample size)

  N - total population size
  M - total number of successes in the population
  n - sample size
  x - successes in sample

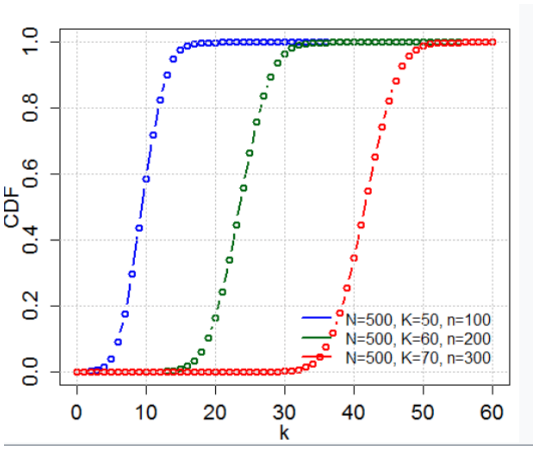
- *What do the PDF and CDF represent, in general terms?*
  The PDF is the discrete probability distribution which tells you the probability of one specific outcome.
  The CDF is the continuous cumulative probability distribution that greater than or less than of a particular outcome will happen.

- *What do the PDF and CDF for this distribution look like, generally speaking?*
  X-axis is sample size while y-axis is probability.
  Pdf - bell curve (in the graphs, r is x and K is M)





  Cdf - looks like logistic



- *How does changing parameter values affect the probability distribution?*
  - increasing sample size (n) increases mean, increases variance in pdf
  - increasing total successes (M or K in images) increases mean, decreases variance
  - increasing total pop (N) decreases mean, decreases variance

  Mean = nM/N

  Variance = $n\left(\frac{M}{N}\right)\left(1-\frac{M}{N}\right)\left(\frac{N-n}{N-1}\right)$

  (variance formula is a little more complicated so may not be exact)

- *What is one other distribution that relates to your distribution(s)? This could be a special case, a limiting condition, an approximation, an inverse function, etc.*

  Geometric looks at first success after x failures whereas the hypergeometric looks at the number of successes in a sample size.

  Binomial looks at successes in a population based on a probability of success, whereas hypergeometric has a finite population that you sample from and looking at probability of so many successes in the sample

  Most directly related is the one tailed Fisher's exact test which is the same as the hypergeometric test.