NOTES FROM MONDAY, MARCH 6

I. (p. 238, Section 3.4)

"Population size doesn't matter: The **variability** of a statistics from a random sample does not depend on the size of the population, as long as the population is at least 100 times larger than the sample."

Analogy: "A scoop of corn doesn't know whether it is surrounded by a bag of corn or by an entire truckload."

This is <u>not</u> saying anything about how representative the sample is.

What it <u>does</u> say is that for larger sample sizes, we have a higher degree of assurance that our estimate (the statistic) is within a smaller range of the parameter we are trying to estimate.

In particular, it is telling us that we can have some degree of assurance of how close our estimate is without knowing whether or not the sample is representative of the whole population.

II. (p. 342, Section 5.1)

The standard deviation of the sampling proportion p^ of successes in an SRS of size n drawn from a large population having population proportion p of successes is

$$\sqrt{\frac{p(1-p)}{n}}$$

The formula for standard deviation is exactly correct in the binomial setting. It is approximately correct for an SRS from a large population.

Rule of thumb: This approximation is reasonable when the population is at least 20 times as large as the sample.

This is not saying anything about how representative the sample is.

What it <u>does</u> say is that for larger sample sizes, we have a higher degree of assurance that our estimate (the statistic) is within a smaller range of the parameter we are trying to estimate.

It also gives us a measure of the degree of assurance -- provided the Rule of Thumb is met.

III. (p. 344, Section 5.1) In the same setting as II, the distribution of p^ is approximately normal.

Rule of thumb: This approximation is reasonable when both np and n(1-p) ≥ 10.

This is not saying anything about how representative the sample is.

What it <u>*does*</u> *say is that for larger sample sizes, we can do calculations based on normal distributions -- provided the Rule of Thumb is met.*

IV. MORE DETAIL: FINITE POPULATION CORRECTION

A. (Counts and proportions)If we take samples of size n *without* replacement, with probability of success p each time, and the population has finite size N, and we find the number of success, the resulting sampling distribution is *hypergeometric* (rather than binomial). The mean of the distribution is still np, but the standard deviation is

$$\sqrt{\frac{N-n}{N-1}}\sqrt{np(1-p)}$$

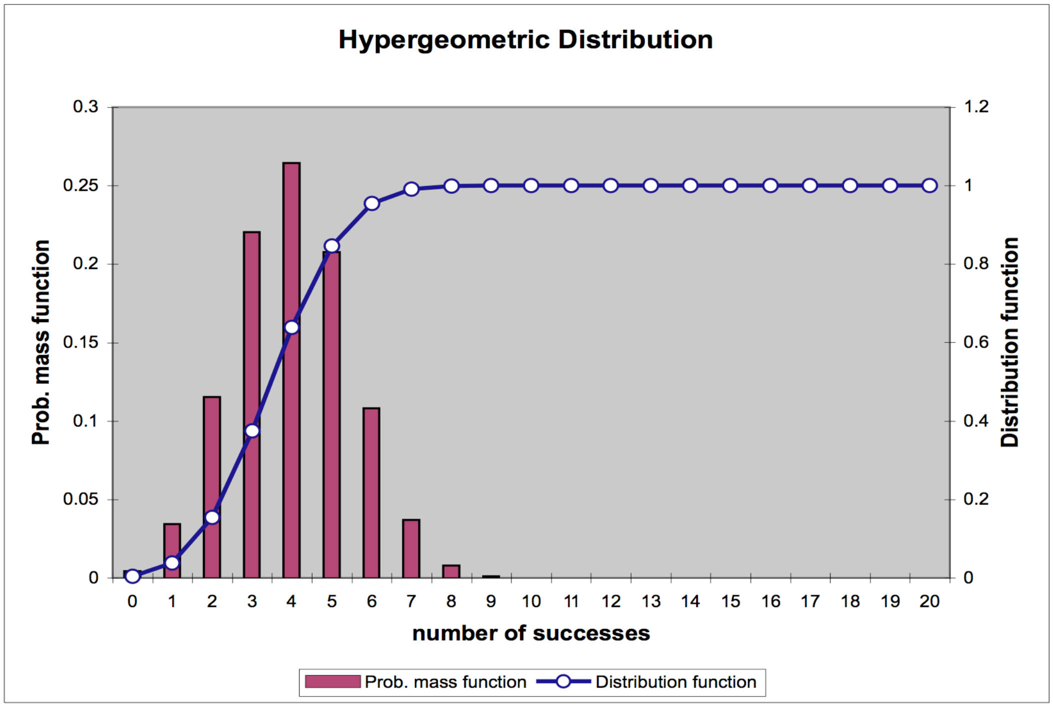
If we look at the distribution of the *proportion* of successes, we get mean p and standard deviation of

$$\sqrt{\frac{N-n}{N-1}}\sqrt{\frac{p(1-p)}{n}}$$

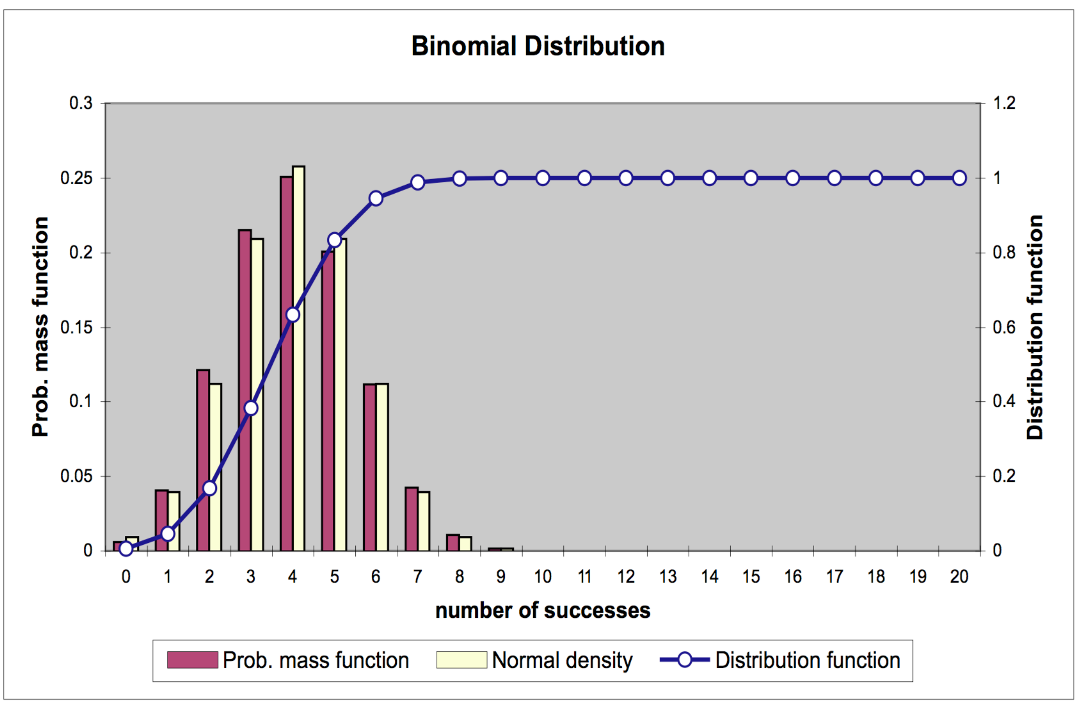
B. (Means) The sampling distribution of means of samples of size n taken *without* replacement from a finite population of size N with mean μ and standard deviation σ has mean μ but standard deviation

$$\sqrt{\frac{N-n}{N}}\frac{\sigma}{\sqrt{n}}$$

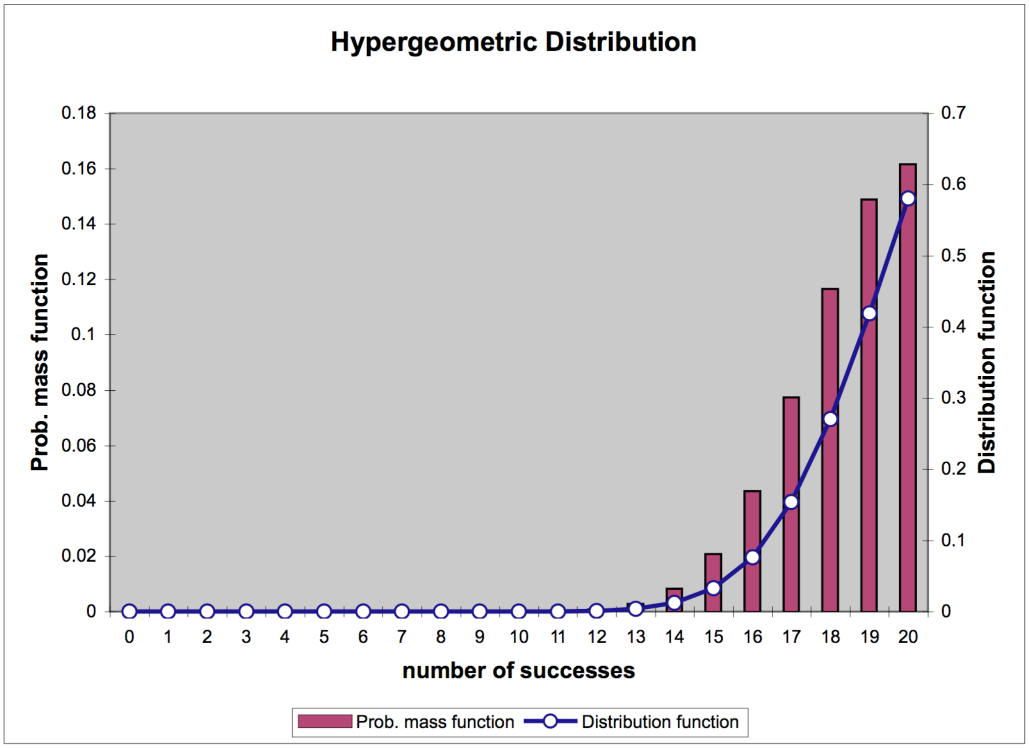
Examples: Comparing hypergeometric, binomial, and normal distributions in two cases



N = 100, p = 0.4, n = 10



(White bars are normal with same mean and standard deviation.)



N = 100, p = 0.4, n = 50 (Just part of distribution) vnbv