Section 3.4: The Hypergeometric Probability Distribution

Tuesday, January 30, 2018   4:01 PM

suppose you are sampling without replacement

remember the rule to approximate with binomial

$\frac{n}{N} < \frac{1}{20}$ to approximate with binomial

why? because you need p to be more or less constant throughout trials to approximate with binomial

but what if you are sampling without replacement and $\frac{n}{N} \geq \frac{1}{20}$?

use hypergeometric probability distribution instead

---

hypergeometric distribution:

population with a total number of N contains M successes and $N-M$ failures

choose without replacement

then the probability of exactly k successes in a random sample of size n is:

$$P(x=k) = \frac{{}_{M}C_{k} \quad {}_{N-M}C_{n-k}}{{}_{N}C_{n}}$$

with mean

$$\mu = n\left(\frac{M}{N}\right)$$

and variance

$$\sigma^2 = n\left(\frac{M}{N}\right)\left(\frac{N-M}{N}\right)\left(\frac{N-n}{N-1}\right)$$

note: this distribution is like the binomial but corrected for a finite population

---

example: A case of wine (12 bottles in total) has 5 bottles that contain spoiled wine. If 3 bottles are randomly sampled, what is the probability distribution for x, the number of spoiled bottles sampled?

- without replacement
- sampling 3 out of 12

so population size is less than 20 times the sample size

$N = 12$

$M = 5$

$N - M = 7$

$n = 3$

$k =$ our random variable X