§3.5 Hypergeometric & Negative Binomial (& Geometric)

Review: Bernoulli(p)

A single measurement: "Bernoulli trial" $X = \begin{cases} 1 & \text{"success"} \\ 0 & \text{"failure"} \end{cases}$

- p = Probability of $X = 1$

Binomial(n, p)

Multiple measurements: n Bernoulli trials X = #success

- n = total number of trials
- p = prob. of each trial being success

New: HyperGeometric (M, N, k)

Measure a "sample" subset chosen from fixed population of known size with known total #successes

Chosen Sample	Total Population
Size = k (#measurements)	# successes = M
# success = X	# not succ. = N

(Note: If you think of this as n Bernoulli trials then the prob (success) is changing after each trial)

Negative Binomial (r, p)

Perform Bernoulli(p) trials until #success = r

- r = #(success required)

X = # failures seen.

- p = probability of success

WARNING: Some people define negative binomial as
X = # trials needed for r successes
instead of
X = # failures before r successes
(Since of #trials is what we are often interested in...)

BUT ① "Negative" in name Negative Binomial is from counting negative results (failures)
② R defines **nbinom** as counting #failures

Extra: Geometric (p)

- Negative Binomial $(1, p)$ counts #failures before 1 succ.
  This is computing "Waiting Time" for success
- Geometric (p) counts # trials needed for 1 succ.
  (If $X \sim$ Geometric(p) then $X - 1 \sim$ Negative Binomial$(1, p)$)

The pmf and cdf for these distributions are rather unpleasant so we use computers.

Distribution	p.m.f. $P(X = x)$
Binomial (n, p)	dbinom (x, n, p)
Hypergeom (N, M, k)	dhyper (x, N, M, k)
Neg. Binom. (r, p)	dnbinom (x, r, p)
Geometric (p)	dgeom (x, p)

p.m.f., mean, and variance (not important for us)

Review: $X \sim$ Bernoulli (p)

$$f(x) = \begin{cases} p & \text{if } x=1 \\ 1-p & \text{if } x=0 \end{cases} \qquad \left[\begin{array}{l} \mu = p \\ \sigma^2 = p(1-p) \end{array}\right.$$

$X \sim$ Binomial (n, p)

$$f(x) = \binom{n}{x} p^x (1-p)^{n-x} \qquad \left[\begin{array}{l} \mu = np \\ \sigma^2 = np(1-p) \end{array}\right.$$

New: $X \sim$ HyperGeom (M, N, k)

(Note: $\frac{M}{N+M} \approx p$)

$$f(x) = \frac{\binom{M}{x}\binom{N}{k-x}}{\binom{N+M}{k}} \qquad \left[\begin{array}{l} \mu = \frac{M}{N+M} \\ \sigma^2 = \underbrace{\left(\frac{N+M-n}{N+M-1}\right)}_{\text{"finite population correction factor"}} k \frac{N}{N+M}\left(1 - \frac{N}{N+M}\right) \end{array}\right.$$

$X \sim$ Neg. Binomial (r, p)

$$f(x) = \binom{x+r-1}{r-1} p^r (1-p)^x \qquad \left[\begin{array}{l} \mu = r \cdot \frac{1}{p} - r = \frac{r(1-p)}{p} \\ \sigma^2 = \frac{r(1-p)}{p^2} \end{array}\right.$$

$X \sim$ Geometric (p)

$$f(x) = p(1-p)^{x-1} \qquad \left[\begin{array}{l} \mu = 1/p \\ \sigma^2 = \frac{1-p}{p^2} \end{array}\right.$$

$$F(x) = 1 - (1-p)^x \quad (!!)$$

Example: Suppose $1/10$ of machine output is flawed.

(I) Pick one output. X = flawed or not?

$X \sim$ Bernoulli $(1/10)$

(II) Pick 20 outputs. X = # flawed

$X \sim$ Binomial $(20, 1/10)$

(III) Pick outputs until 20 good. X = # flawed

$X \sim$ Neg. Binom $(20, 9/10)$

(IV) Pick outputs until first flaw. X = # outputs

$X \sim$ Geometric $(1/10)$

(V) Collect box of 50 outputs, $1/10$ of which are flawed. Pick 8 from box. X = # flawed

$X \sim$ HyperGeom $(5, 45, 8)$

N: #(flawed) M: #(good) k: #(picked)