Lecture 10 10/10/2017

10 cards, 4R, 6B

$$P(2 \text{ Reds in } 3 \text{ cards}) = \frac{\binom{4}{2}\binom{6}{1}}{\binom{10}{3}}$$

$$P(x R \text{ in } 3 \text{ cards}) = \frac{\binom{4}{x}\binom{6}{3-x}}{\binom{10}{3}}$$

$$P(x R \text{ in } n \text{ cards}) = \frac{\binom{4}{x}\binom{6}{n-x}}{\binom{10}{n}}$$

K reds, 10-K blues $\quad P(x R \text{ in } n \text{ cards}) = \frac{\binom{K}{x}\binom{10-K}{n-x}}{\binom{10}{n}}$

N cards $\quad P(x R \text{ in } n) = \frac{\binom{K}{x}\binom{N-K}{n-x}}{\binom{N}{n}}$

$$\underset{\text{r.v distributed as}}{X \sim} \text{HyperGeometric}\underbrace{(n, K, N)}_{p(x) \text{ PMF}} := \overbrace{\frac{\binom{K}{x}\binom{N-K}{n-x}}{\binom{N}{n}}}^{p(x)}$$ without replacement

$X \sim \text{Bern}(p) := p^x(1-p)^{1-x}$ Shorthand notation for PMF

$\underset{\text{possibilities}}{\text{Supp}[x]} = \{0,1\} \quad \underset{\text{parameter space}}{p \in (0,1)}$

100 "thumps"

Hypogeometric → pulling arbitrary amount

~~Hyperbeometric (n, K, N)~~: $\frac{\binom{K}{x}\binom{N-K}{n-x}}{\binom{N}{n}}$

Supp[X] = ?

100 students
53 are Females
pick 8 at random
provide X, a r.v model which counts the # of Females

$$X \sim \text{Hypogeometric}(8, 53, 100)$$

$$= \frac{\binom{53}{x}\binom{100-53}{8-x}}{\binom{100}{8}} = p(x)$$

$\binom{100-53}{8-x} = \binom{47}{8-x}$

x = free variable

probability of x of 8 are females

Parameter Space of the Hypogeometric R.V

$N = 0$? "no marbles" No!
$n = 0$? "picking no marbles" No!

$N = 1 \Rightarrow K \in \{0, 1\} \Rightarrow n = 1$

if $K = 0$ degenerate! $X \sim Deg(0)$
if $K = 1$ Degenerate! $X \sim Deg(1)$

$N = 2$ $n \neq 0$ $n = 2$? $X \sim Deg(K)$

$n = 1$ $K = 0 \Rightarrow X \sim Deg(0)$

Success marbles

$K = 2 \Rightarrow X \sim Deg(1)$

Bag of 2 marbles, pull one out

$K = 1 \Rightarrow X \sim \text{Hypogeometric}(1, 1, 2) = \frac{\binom{1}{x}\binom{2-1}{1-x}}{\binom{2}{1}} = Bern\left(\frac{1}{2}\right)$

chance of failure

$p(0) = \frac{1 \cdot 1}{2} = \frac{1}{2}$

$p(1) = \frac{1 \cdot 1}{2} = \frac{1}{2}$

$$X \sim Hyper(1, K, N) = \frac{\binom{K}{x}\binom{N-K}{1-x}}{\underbrace{\binom{N}{1}}_{N}} = Bern\left(\frac{\overset{\#\text{ of success}}{K}}{N}\right)$$

$$P(0) = \frac{\binom{K}{0}\binom{N-K}{1}}{N} = \frac{N-K}{N} = 1 - \frac{K}{N} = \text{"}1-p\text{"}$$

$$p(1) = \frac{\binom{K}{1}\binom{N-K}{0}}{N} = \frac{K}{N} = \text{"}p\text{"}$$

Proof for Bernoulli

$$\left.\begin{array}{l} N \in \{2,3,4 \ldots\} \\ K \in \{1,2, \ldots N-1\} \\ n \in \{1,2 \ldots N-1\} \end{array}\right\} \text{Parameter space}$$

$Supp[X] = \{0,1\}$ => all possible values / realizations

a) $X \sim Hyper(2,4,10)$, $Supp[X] = \{0,1,2\}$
"Bag of 10 marbles, where there are 4 possible successes, how many of 2 are successes, how many failures w.o replacement

b) $X \sim Hyper(5,4,10)$, $Supp[X] = \{0,1,2,3,4\}$
c) $X \sim Hyper(8,4,10)$, $Supp[X] = \{2,3,4\}$
d) $X \sim Hyper(5,7,10)$, $Supp[X] = \{2,3,4,5\}$

NOTE: pick 1, will have 4 failures impossible can only have 3 failures

a) $n < K$, $n < \overset{\text{total \# of failures}}{N-K}$ $\quad Supp[X] = \{0,1, \ldots n\}$

b) $n > K$, $n < N-K$ $\quad Supp[X] = \{0,1, \ldots K\}$

c) $n \geq K,\ n \geq N-K \quad Supp(X) = \{n-(N-K), \dots K\}$

d) $n < K,\ n \geq N-K \quad Supp(X) = \{n-(N-K) \dots n\}$

	$n<K$	$n \geq K$
$n<N-K$	$\{0..n\}$	$\{0..K\}$
$n \geq N-K$	$\{n-(N-K) \dots n\}$	$\{n-(N-K) \dots K\}$

$$\Downarrow$$

$$supp(X) = \{\max\{0, n-(N-K)\}, \dots \min\{n, K\}\}$$

$$\sum_{x \in supp(X)} p(x) = 1 \qquad \sum_{x \in supp(X)} \frac{\binom{K}{x}\binom{N-K}{n-x}}{\binom{N}{n}} = 1$$

Equivalent Parameterization

let $p = \frac{K}{N} \therefore K = pN$

$X \sim Hyper(n, p, N)$

$$PMF \quad \frac{\binom{pN}{x}\binom{\overset{(1-p)N}{N-(pN)}}{n-x}}{\binom{N}{n}}$$

$$p \in \{\frac{1}{N}, \frac{2}{N}, \dots \frac{N-1}{N}\}$$

$p = 0.5, n=6, N=100 \qquad p(3) = \frac{\binom{50}{3}\binom{50}{3}}{\binom{100}{6}} = .3223$

$N = 1000 \quad p(3) = \frac{\binom{500}{3}\binom{500}{3}}{\binom{1000}{6}} = .3134$

$N = 10000 \quad p(3) = \frac{\binom{5000}{3}\binom{5000}{3}}{\binom{10000}{3}} = .3126$

$\therefore \lim_{N\to\infty} P(X) = \dots$ closer to sampling w/ replacement