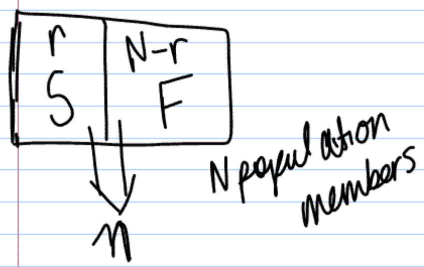
Hypergeometric

Note Title 10/10/2008



Y = # of successes in n draws

in n trials

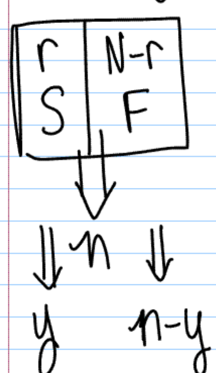
in the sample of size n

$$P(Y=y) = \frac{\binom{r}{y}\binom{N-r}{n-y}}{\binom{N}{n}}$$

$y = 0, 1, 2, \ldots, n$

restricted to

$y \le r$ & $n-y \le N-r$



$$\underline{F} \quad \underline{F} \quad \underline{F} \quad \underline{F} \quad \underline{F} \qquad n=5$$

$$\frac{N-r}{N} * \frac{N-r-1}{N-1} * \frac{N-r-2}{N-2} * \frac{N-r-3}{N-3} * \frac{N-r-4}{N-4} \qquad y=0$$

$$= \frac{\binom{N-r}{5}}{\binom{N}{5}}$$

$$\underline{S} \quad \underline{F} \quad \underline{F} \quad \underline{F} \quad \underline{F} \qquad n=5$$

$$\frac{r}{N} * \frac{N-r}{N-1} * \frac{N-r-1}{N-2} * \frac{N-r-2}{N-3} * \frac{N-r-3}{N-4} \qquad y=1$$

$$\underline{F} \quad \underline{F} \quad \underline{S} \quad \underline{F} \quad \underline{F}$$

$$\frac{N-r}{N} * \frac{N-r-1}{N-1} * \frac{r}{N-2} * \frac{N-r-2}{N-3} * \frac{N-r-3}{N-4}$$

$$\frac{\binom{r}{1}\binom{N-r}{4}}{\binom{N}{5}}$$

Mean & variance

$\mu = E(Y) = n\left(\frac{r}{N}\right)$

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

$\frac{r}{N} = p =$ probability of success in population.

$\mu = np$

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

$\left(\frac{N-n}{N-1}\right)$ – correction for finite population

In most cases, n is small compared to N

So, even though we are sampling without replacement, the statistics are based on the binomial dist.

F	S
8	92

$\downarrow$

5

3.

15	25
S	F

40

Y = # of yellow in sample of 3

(a) $P(Y=3) = 0.0461 = \dfrac{\binom{15}{3}\binom{25}{0}}{\binom{40}{3}}$

(b) $P(Y=2) = 0.2657$

(c) $\mu = n\left(\frac{r}{N}\right) = 3 \cdot \frac{15}{40} =$

(d) $\sigma^2 = n\left(\frac{r}{N}\right)\left(\frac{N-r}{N}\right)\left(\frac{N-n}{N-1}\right) = 3 \cdot \left(\frac{15}{40}\right)\left(\frac{25}{40}\right)\left(\frac{37}{39}\right)$

(4.)

10	110
BC	No BC
S	F

N=120

3

(a) $P(Y=1)$

(b) Set r to be between 1 + 120.

```
r<- 1:120
```

Then calculate $P(Y=1)$ for all r.

```
probs1<- dhyper(1, r, 120-r, 3)
```

c)

```
plot(1:120, dhyper(1, r, 120-r, 3))
```

You can see in the plot, the probability is maximized when r = 40

(5.) $N = 240$, $n = 60$, $r = 14$.

(a) $P(Y=6) = 0.0713$

(b) $P(0 \le Y \le 14) =$ dhyper(0:14, 14, 240-14, 60)

(c) Yes, this would not happen very often by chance.

The probability $P(Y=14) = 1.0580 \times 10^{-9}$

(d) $100 * P(Y=6) = 7.13$

(e) $P(Y=14) = 1.0580 \times 10^{-9}$. The number of shuffles S needed to hear all 14 Queen songs just once is a geometric dist. $E(S) = \frac{1}{1.0580 \times 10^{-9}} = 945{,}159{,}100$