* Announcement
① HW3 due next Tue (9/15)
② Office hour structure - waiting room
③ Study group survey (around this weekend)

STAT 88: Lecture 7

Contents

Section 3.4: The Hypergeometric Distribution
Section 3.5: Examples

Last time

Sec 3.3 The binomial distribution has 2 parameters, Binomial(n, p):

- $n = \#$ independent trials
- $p =$ probability of success
- $X = \#$ successes out of n trials $\sim$ Binom(n,p)

Binomial formula:

$$P(X = k) = \binom{n}{k} p^k (1-p)^{n-k}. \longleftarrow \text{Generalize.}$$

$k = 0, 1, 2, \cdots, n$

$\binom{5}{3} \cdot p^3(1-p)^2$

$\Uparrow$

$n=5$ trials
Head = success
$p = P(\text{Head})$

$\underline{H}\ \underline{H}\ \underline{H}\ \underline{T}\ \underline{T} \rightsquigarrow p^3(1-p)^2$

There are $\binom{5}{3}$ many permutations

Warm up: 13 cards are dealt from a deck with replacement:

(a) Find the chance that the hand contains two aces.

$k=2$

(b) Find the chance that the hand contains more than two aces.

$k>2$

(c) Find the chance that the hand contains six face cards.

$k=6$

(a) X = # aces in the hand.

$X \sim \text{Binomial}(13, \frac{4}{52})$

$P(X=2) = \binom{13}{2} * \left(\frac{4}{52}\right)^2 \left(\frac{48}{52}\right)^{11}$

(b) $P(X>2) = P(X=3) + P(X=4) + \cdots + P(X=13)$

$1-P(X\le 2)$ $\quad = \sum_{k=3}^{13} \binom{13}{k} * \left(\frac{4}{52}\right)^k \left(\frac{48}{52}\right)^{13-k}$

(c) X = # face cards in the hand (J, Q, K)

$X \sim \text{Binomial}(13, \frac{12}{52})$

$P(X=6) = \binom{13}{6} * \left(\frac{12}{52}\right)^6 \left(\frac{40}{52}\right)^7.$

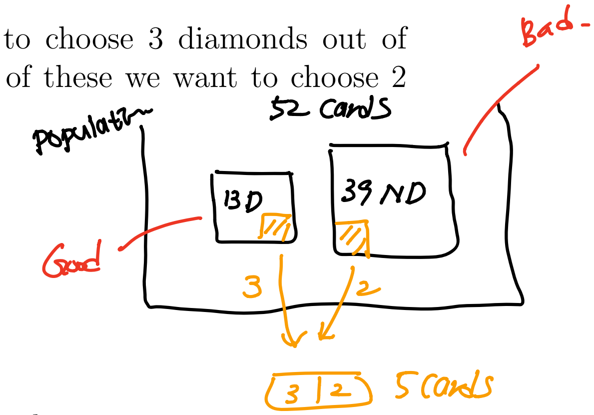
3.4. The Hypergeometric Distribution

When you are sampling at random from a finite population, it is more natural to draw without replacement than with replacement.

Hyper Geom   Binom.

Example: Five cards are dealt at the top of a deck. Find the chance of getting exactly 3 diamonds.

Let X = # diamonds out of 5 cards. We want to choose 3 diamonds out of 13(=52/4). There are $\binom{13}{3}$ ways to do this. For each of these we want to choose 2 nondiamonds out of 39 $\longrightarrow \binom{39}{2}$.



Since all $\binom{52}{5}$ sample are equally likely we get

$$P(X=3) = \frac{\binom{13}{3}\binom{39}{2}}{\binom{52}{5}}.$$

More generally the ingredients of a hypergeometric distribution are:

- N = population size (52 Card deck)
- G = # good elements in your population ($B = N - G$ is the number of bad elements) (13 ~~aces~~ Diamonds) (39 non ~~aces~~ Diamonds.)
- n = # sample size (5 cards)

Let X = # good elements in your sample. Then the hypergeometric formula is

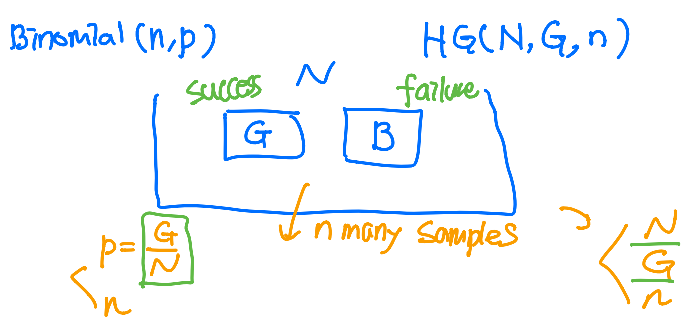
(3 cards)

$$P(X=g) = \frac{\binom{G}{g}\binom{B}{b}}{\binom{N}{n}}.$$

B= N-G
b= n-g

g= 0, 1, ---, min(n, G)

We say $X \sim \mathrm{HG}(N, G, n)$.

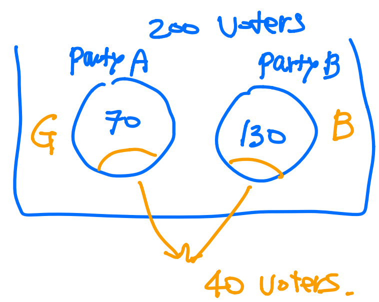


Example: (Exercise 3.6.6) In a population of 200 voters, 70 are registered with Party A and the other 130 are registered with Party B. A simple random sample of 40 voters is drawn from this population. Let $\cancel{V}$ X be the number of sampled voters who are registered with Party A, and let $W = 40 - \cancel{V}$ X be the number of sampled voters who are registered with Party B. Find:

(a) $P(V = 10)$

(b) $P(V > 10)$

(c) $P(W < 3V)$



(a) $P(X=10) = \dfrac{\binom{70}{10}\binom{130}{30}}{\binom{200}{40}}$

$N = 200$

$n = 40$

$G = 70.$

$X =$ # sampled voters w/ party A.

(b) $P(X > 10) = P(X=11) + P(X=12) + \cdots + P(X=40)$

$= \sum_{k=11}^{40} \dfrac{\binom{70}{k}\binom{130}{40-k}}{\binom{200}{40}}$

(c) $P(W < 3X) = P(40 - X < 3X)$

$= P(X > 10)$

$=$ Same as part (b)

Hypergeometric Probabilities in Python You can use the stats module of SciPy to calculate hypergeometric probabilities, just as you used it to calculate binomial probabilities.

Hypergeometric formula:

$$P(X = g) = \frac{\binom{G}{g}\binom{B}{b}}{\binom{N}{n}}.$$

```
from scipy import stats
import numpy as np
```

$$\frac{\binom{70}{10}\binom{130}{30}}{\binom{200}{40}} \rightarrow$$

g N G n

```
stats.hypergeom.pmf(10, 200, 70, 40)
```

```
0.05054861360578296
```

```
sum(stats.hypergeom.pmf(np.arange(11,71), 200, 70, 40))
```

```
0.9043345335065547
```

$$\sum_{g=11}^{70} \frac{\binom{70}{g}\binom{130}{40-g}}{\binom{200}{40}}$$

min(n,G)
''

g can only take values from 0, 1, 2, ..., 40.
If $g \geq 41$, then $40-g$ in $\binom{130}{40-g}$ becomes negative.

In textbook and Python, they extend the definition that quantities like $\binom{130}{-2}$ and $\binom{70}{75}$ are "0". i.e. there is no way to choose -2 elements out of 130 and there is no way to choose 75 elements out of 70.