stat 88 lec 7

warm up 2:00–2:10

e.g. 13 cards are dealt from a deck with replacement

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

$\binom{13}{2}\left(\frac{1}{13}\right)^2\left(\frac{12}{13}\right)^{11}$

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

$\sum_{k=3}^{13}\binom{13}{k}\left(\frac{1}{13}\right)^k\left(\frac{12}{13}\right)^{13-k}$

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

$\binom{13}{6}\left(\frac{3}{13}\right)^6\left(\frac{10}{13}\right)^7$

$\frac{12}{52}$

last time

Sec 3.3 Binomial Distribution,

The binomial distribution has 2 parameters

Binomial (n, p)

n = # independent trials

p = probability of success,

X = # successes out of n trials

$$P(X=k) = \binom{n}{k} p^k (1-p)^{n-k}$$ binomial formula

H H H T T $p^3(1-p)^2$

a b c d e $\frac{5!}{3!\,2!} = \binom{5}{3}$

Today ① review concept test from last time

② sec 3.4 The hypergeometric distribution

③ sec 3.5 More examples

Discuss

Ten cards are dealt off the top of a well shuffled deck. The binominal formula doesn't apply to find the chance of getting exactly three diamonds because:

a The probability of a trial being successful changes

b The trials aren't independent

c There isn't a fixed number of trials

d more than one of the above

(1) review of current test

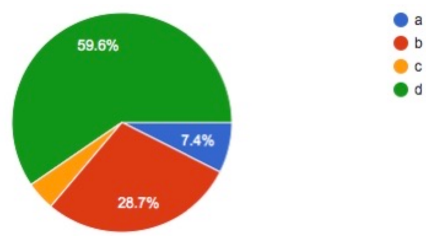
Ten cards are dealt off the top of a well shuffled deck. The binominal formula doesn't apply to find the chance of getting exactly three diamonds because:

a The probability of a trial being successful changes

b The trials aren't independent

c There isn't a fixed number of trials

d more than one of the above



d	a and b are linked

d	Every time you take a card out of the deck, it changes the probability of what is left in the deck, and an individuals chance of being a diamond.

b	The unconditional probability of getting a diamond is always 1/4. If we were drawing from two decks of cards having different number of diamonds than a would be false.

(2) Sec 3.4 <u>hypergeometric distribution</u>

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

ex five cards are dealt of 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. 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 $\binom{52}{5}$ samples are equally likely we get

$$P(X=3) = \frac{\overset{D}{\binom{13}{3}}\overset{N}{\binom{39}{2}}}{\underset{n}{\binom{\overset{N}{52}}{5}}}$$

hypergeometric formula

more generally the ingredients of a hypergeometric distribution are

N = population size (52 card deck)

G = # good element in your population (13 aces)

($B = N - G$ is the number of bad elements) (39 nonaces)

n = sample size (5 cards)

let X = # good elements in your sample (3 aces)

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

we say $X \sim HG(N, G, n)$ (X belongs to the hypergeometric dist.)

ex (exercise 3.6.6)

Biden major Pete

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 X be the number of sampled voters who are registered with Party A, and let $b = 40 - X$ be the number of sampled voters who are registered with Party B Find:

a) $P(X = 10)$ $\quad \dfrac{\overset{G}{\binom{70}{10}}\overset{B}{\binom{130}{30}}}{\binom{200}{40}}$ $\quad N = 200$, $G = 70$, $n = 40$

b) $P(X > 10)$ $\quad \displaystyle\sum_{g=11}^{40} \frac{\binom{70}{g}\binom{130}{40-g}}{\binom{200}{40}}$

c) $P(\overset{40-X}{b} < 3X) = P(40-X < 3X) = P(40 < 4X) = P(X > 10)$

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

Hypergeometric probabilities in Python:

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

hypergeometric formula

$$\frac{\overset{G}{\binom{70}{10}}\overset{B}{\binom{130}{30}}}{\binom{200}{40}}$$

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

```
In [5]: from scipy import stats
        import numpy as np
```

$g \quad N \quad G \quad n$

```
In [6]: stats.hypergeom.pmf(10,200,70,40)
Out[6]: 0.05054861360578296
```

```
In [9]: sum(stats.hypergeom.pmf(np.arange(11,71), 200, 70, 40))
Out[9]: 0.9043345335065547
```

(3) Sec 3.5 More examples

Problem solving techniques:

- organize the info to identify parameters
- Partition events into component pieces.
- use addition and multiplication rules.

Randomized Controlled Experiment (RCE)

In a RCE a SRS of half the participants are assigned to a treatment group (T) and half to a control group (C).

Experiment 1 has 100 participants of whom 20 are men.

What is the chance that T and C have the same number of men?

Since SRS use HG(N, G, n)

— organize the info to identify parameters

sample is treatment group

$N = 100$

$n = 50$

$G = 20$

$P(X = \cancel{9}^{10})$

X = # of good in your sample

good = men; sample = treatment group

$$P(X=10) = \boxed{\frac{\overset{G}{\binom{20}{10}}\overset{B}{\binom{80}{40}}}{\binom{100}{50}}}$$

Experiment 1 has 100 participants of whom 20 are men.

Experiment 2 has 90 participants of whom 30 are men

What is the chance ~~that~~ the treatment groups in the two experiments have the same number of men?

- organize the info to identify parameters
- Partition events into component pieces.
- use addition and multiplication rules.

$$P(X_1 = g \text{ and } X_2 = g) = P(X_1 = g)P(X_2 = g)$$

$$= \sum_{g=0}^{20} \frac{\binom{20}{g}\binom{80}{50-g}}{\binom{100}{50}} \cdot \frac{\binom{30}{g}\binom{60}{45-g}}{\binom{90}{45}}$$

Ex (Fisher Exact Test)

A RCE has 100 participants

T=60 C=40

In T, 50 recover out of 60 – 83%

In C, 30 recover out of 40 – 75%

A total of 80 participants recovered out of 100.

Question Suppose the treatment isn't effective.

What is the chance that 50 or more of the recovered patients are randomly assigned to the treatment group?

(if the answer is really small then the treatment is probably effective).

Start with:

What is the chance that 50 of the recovered patients are randomly assigned to the treatment group?

we will continue this next time.