stat 88 lec 25

warmup 2:00-2:10

Draw 6 cards and count the number of red cards you get. To increase your odds of getting 3 red cards should you draw with or without replacement?

Answer: without replacement.

X = # red cards out of 6

$FPC = \sqrt{\frac{N-n}{N-1}} = \sqrt{\frac{52-6}{52-1}} = .95$

$SD(HG) = SD(\text{Binomial}) \, FPC$

so $SD(HG) < SD(\text{Binomial})$

w/ replacement: $X \sim Bin(6, \frac{1}{2})$

$P(X=3) = \binom{6}{3}\left(\frac{1}{2}\right)^6 = \boxed{.313}$

w/o replacement: $X \sim HG(N=52, G=26, n=6)$

$P(X=3) = \frac{\binom{26}{3}\binom{26}{3}}{\binom{52}{6}} = \boxed{.332}$ ← bigger!

Announcement Quiz 3 will be online, Friday April 3 in class covering Chap 6, 7.

Last time

Sec 7.2

$$SD(HG(N,G,n)) = \sqrt{n\frac{G}{N}\frac{N-G}{N}} \cdot \boxed{\sqrt{\frac{N-n}{N-1}}}$$

≤ 1

Finite Population Correction (FPC)

$$SD(HG) = SD(Binomial)\cdot FPC$$

$$\Rightarrow SD(HG) \leq SD(Binomial)$$

Sec 7.3

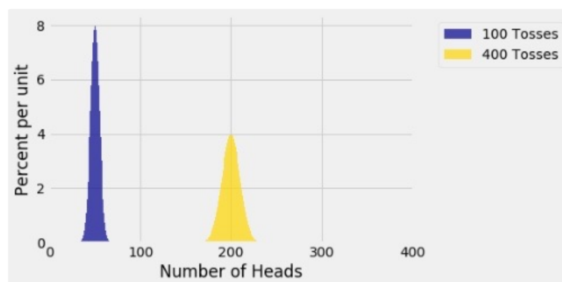
SD of sample sum

$X_1, \dots, X_n \overset{iid}{\sim}$ with mean μ, SD σ

$$S_n = X_1 + \dots + X_n$$

$$E(S_n) = n\mu$$

$$SD(S_n) = \sqrt{n}\,\sigma$$

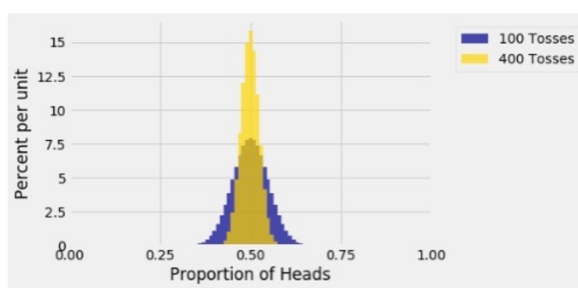


SD of sample average

$$A_n = S_n/n$$

$$E(A_n) = \mu$$

$$SD(A_n) = \frac{\sigma}{\sqrt{n}}$$



Today
1. Concept test from last time
2. Sec 7.3 law of averages
3. Sec 8.1, 8.2 Central limit theorem

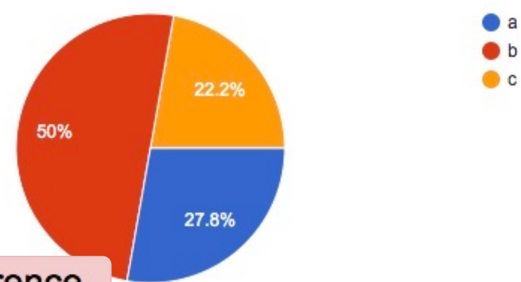
① Concept test from last time

New Mexico has a population of 1M and California a population of 40M. The two states have the same proportion of Democrats. A random sample of size 0.01% of the population is taken. The SD for the number of democrats in the sample is:

a roughly the same in both states

b larger in California

c larger in New Mexico



a	Although there is some difference, since the sample size is so small relative to the population size, the FPC is roughly 1 for both samples
c	LAW OF AVERAGES
b	In percentage ~~they could be the same but~~ and in number of democrats it's different because their total population size is different.

FPC ≈ 1 for both states

σ same both states

n for NM smaller than n for CA

SD(sum) $= \sqrt{n}\,\sigma$ so want state with smaller n

② Sec 7.3 <u>law of averages</u>

You toss a fair coin n times

Let $X_1, \dots, X_n \overset{iid}{\sim} \text{Bernoulli}(\frac{1}{2})$

$S_n = X_1 + X_2 + \dots + X_n$ is the sample sum

let $A_n = S_n/n$ be the sample average

where do you put your money?

a) $A_{100} = 50$ or

b) $A_{400} = 200$?

$SD(\text{Bernoulli}(\frac{1}{2})) = \sqrt{\frac{1}{2}\cdot\frac{1}{2}} = 1/2$

$$SD(A_{400}) = \frac{\frac{1}{2}}{\sqrt{400}} = \frac{1}{40}$$

so A_{400} is twice as accurate as A_{100}

$$SD(A_{100}) = \frac{1/2}{\sqrt{100}} = \frac{1}{20}$$

Conclusion: if you quadruple n, you double the accuracy. This is the square root law.

Square root law says if you multiply the sample size by a factor, the accuracy only goes up by the square root of the factor.

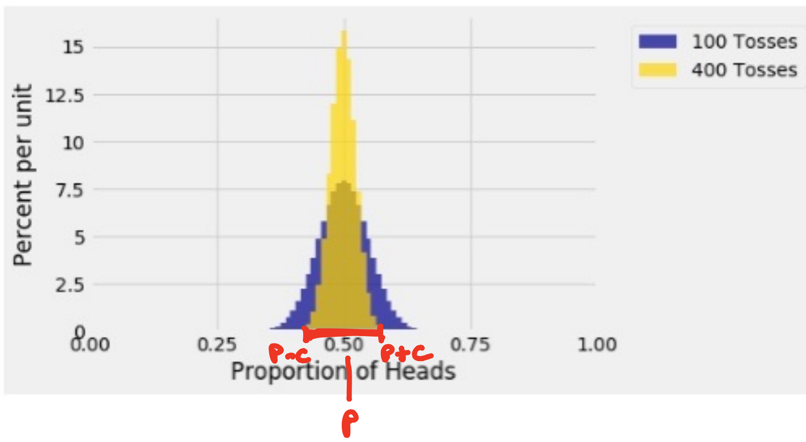
As you increase n to infinity, $SD(A_n)$ goes to zero.

Another way to say this (the law of averages) is:

Let $X_1, \ldots, X_n \overset{iid}{\sim}$ Bernoulli (p)

Let $c > 0$ be any number

Then $P(p - c < A_n < p + c) \to 1$ as $n \to \infty$



(3) Sec 8.1 Distribution of a sample sum

ex exercise 7.4.11

11. Each Data 8 student is asked to draw a random sample and estimate a parameter using a method that has chance 95% of resulting in a good estimate.

Suppose there are 1300 students in Data 8. Let X be the number of students who get a good estimate. Assume that all the students' samples are independent of each other.

a) Find the distribution of X. $X \sim \text{Binomial}(1300, .95)$

b) Find $E(X)$ and $SD(X)$.

$E(X) = 1300(.95)$

$SD(X) = \sqrt{1300(.95)(.05)}$

c) Find the chance that more than 1250 students get a good estimate.

$$\rightarrow \quad P(X > 1250) = \sum_{i=1251}^{1300} \binom{1300}{i} (.95)^i (.05)^{1300-i}$$

In the previous example,

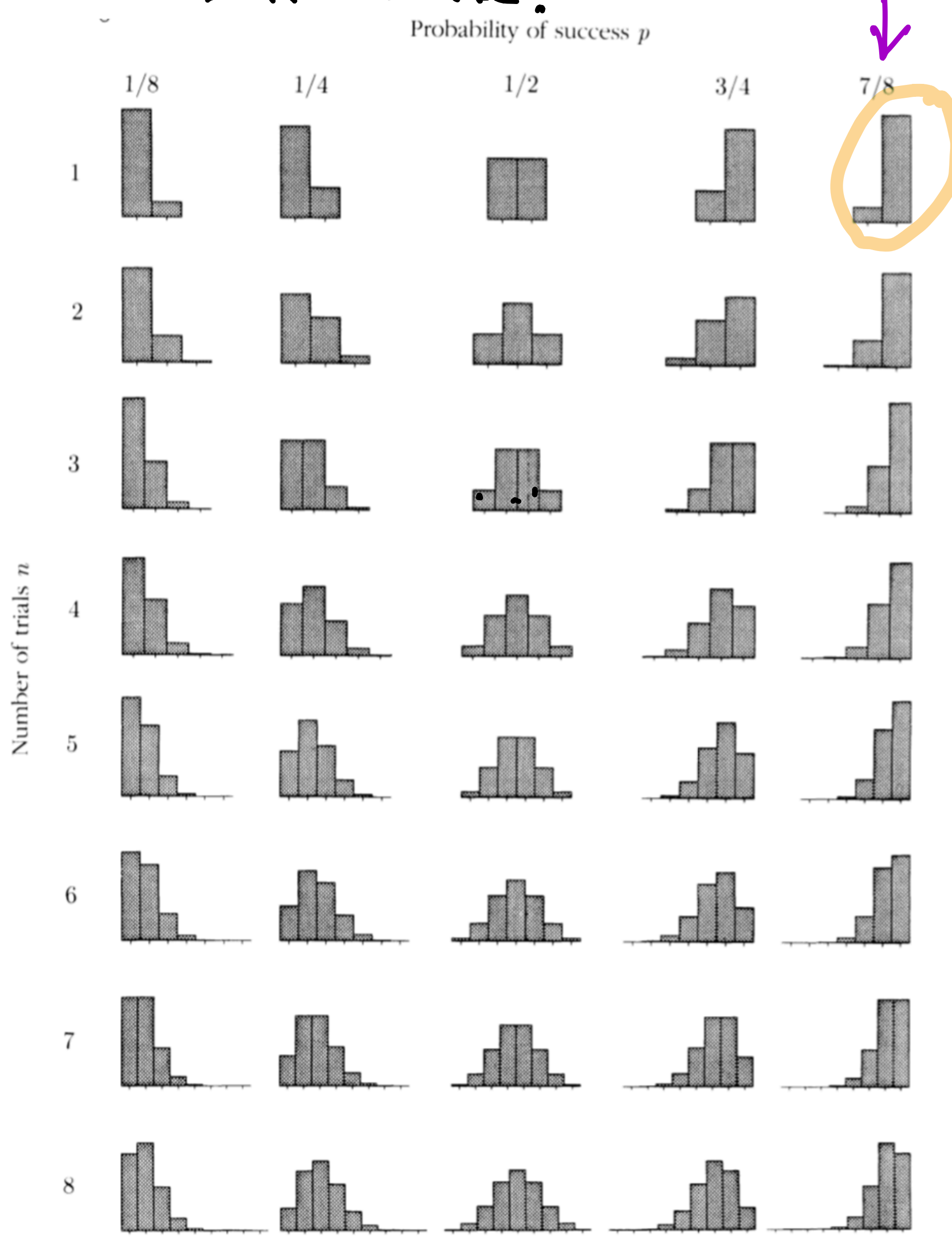
$$X_1, \ldots, X_{1300} \overset{iid}{\sim} \text{Bernoulli}(.95)$$

$$X = X_1 + X_2 + \cdots + X_{1300}$$ (total number of students with a good result).

$$X \sim \text{Binomial}(1300, .95)$$

What does X look like?

close to .95



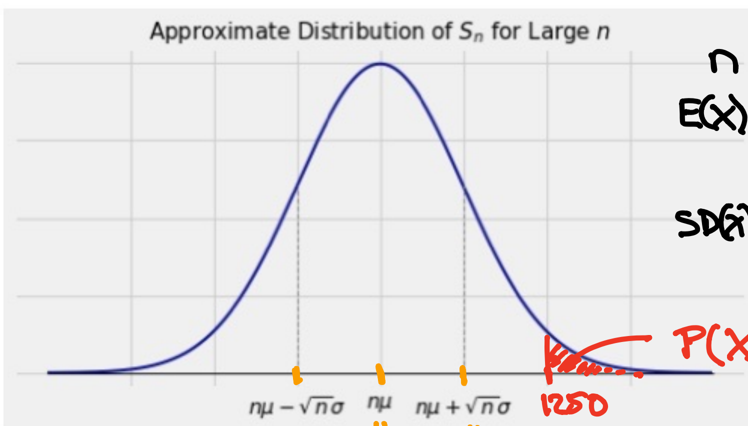
We see that $\bar{X}$ looks like a bell curve even though $X_1, X_2, \ldots, X_n$ dont look like a bell curve.

def^n A bell curve resulting from a large random sample sum is said be normal (i.e a normal curve).

<u>Central Limit Theorem</u> (CLT)

The CLT says:

Let $X_1, \ldots, X_n$ be iid with $E(X_1) = \mu$ and $SD(X_1) = \sigma$. Let $S_n = X_1 + \ldots + X_n$ be the sample sum. If n is large, the distribution of S_n is approximately normal, regardless of the common distribution of the X_i's.



$n = 1300$

$E(X) = 1300(.95)$
$= 1235$

$SD(X) = \sqrt{1300}\sqrt{(.95)(.05)}$
$= 7.86$

$P(X > 1250)$

$= 1227.14 = (1235 - 7.84)$

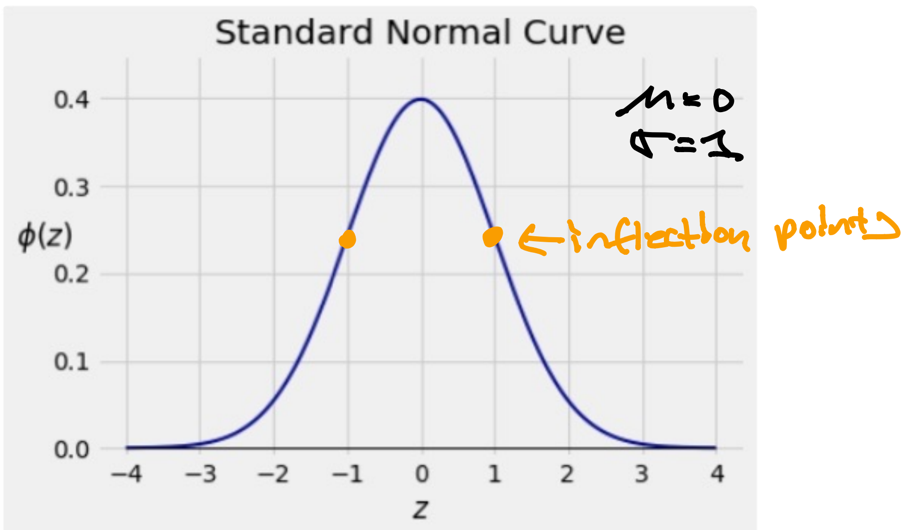
$= 1235$

$= 1242.86$

Key idea: It is easier to approximate $P(X > 1250)$ using the fact that Binomial is almost normal for large n.

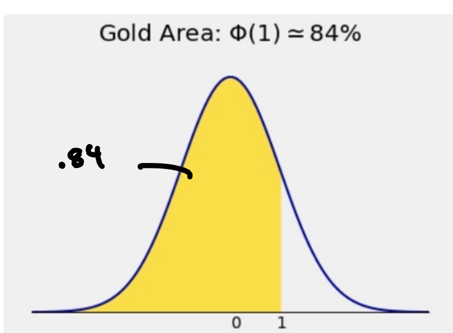
Sec 8.2 Standard normal curve $\phi(z)$

$$\phi(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2}, \qquad -\infty < z < \infty$$



The standard normal CDF $\Phi(z)$

$$\Phi(z) = \int_{-\infty}^{z} \phi(x)\,dx$$



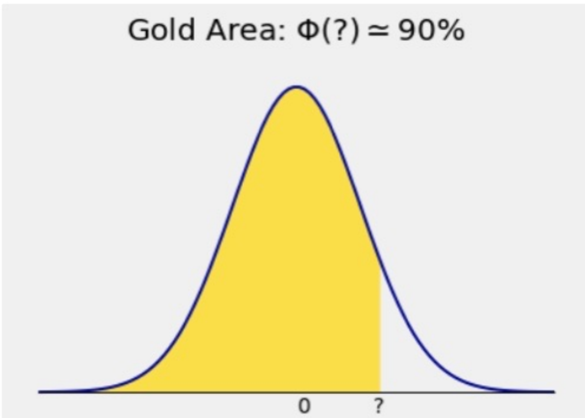
Calculating $\Phi(z)$ in Python:

in SciPy library

```
stats.norm.cdf(1)
```

```
0.8413447460685429
```

Percentiles



$z = \Phi^{-1}(.9)$

Percent Point function

```
z= stats.norm.ppf(0.9)
```

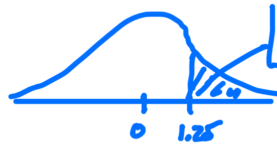
```
1.2815515655446004
```

↖ we need to find the inverse of $\Phi(z)$

The 90^{th} percentile is z such that $\Phi(z) = .9$

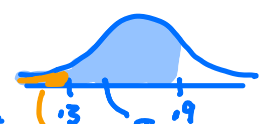
ex Find the area

a) to the right of 1.25



$\boxed{1 - \Phi(1.25)}$

b) between −0.3 and 0.90 — $\Phi(.90) - \Phi(-.3)$

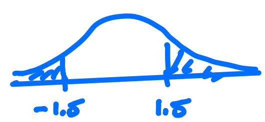


$\Phi(.3)$ $\Phi(.9)$

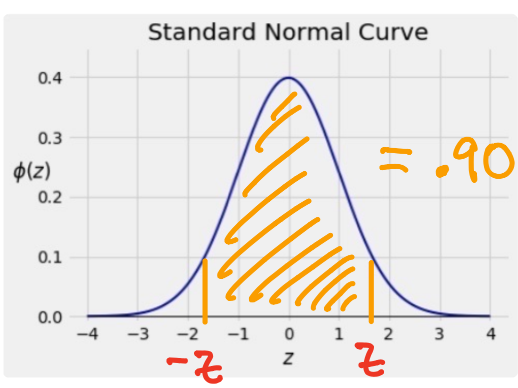
c) outside −1.5 and 1.5

$\Phi(-1.5) + (1 - \Phi(1.5))$

$= 2\Phi(-1.5)$



ex The std normal curve is sketched below,
Solve for z



$\Phi(-z) = .05$

$-z = \Phi^{-1}(.05)$

$\boxed{z = -\Phi^{-1}(.05)}$