What size of variation should we expect?

$P(H) = 1/2$

$P(T) = 1/2$

toss the coin repeatedly, and consider

$$\# \text{heads} - \frac{1}{2} \# \text{tosses}$$

we expect this to be zero

but there will be some variation

- absolute size of the variation will get bigger as # tosses increases.

but the size of the variation as a fraction of the # of tosses goes down.

Toss coin 100 times - record size of the variation
10,000

The # of tosses has increased by a factor of 100, but the size of the variation will only increase by a factor of $\sqrt{100} = 10$

$$\frac{\text{size of variation}}{\#\text{ tosses.}} \longrightarrow \frac{\sqrt{N}}{N}$$

$\Rightarrow$ Size of variation relative to # tosses gets smaller (as $\frac{1}{\sqrt{N}}$).

How do we study problems like these?

$\boxed{0}$ $\boxed{1}$

make an analogy between the process being studied, and drawing numbers at random from a box.

Connect the variability you want to know about with the variability in the <u>sum</u> of the numbers drawn (with replacement) from the box.

<u>Examples.</u>

heads $\boxed{0}$ $\boxed{1}$

1 corresponds to H.

sum of draws is # heads.

$\boxed{0}\ \boxed{0}\ \boxed{0}\ \boxed{0}\ \ldots\ \boxed{1}\ \boxed{1}\ \boxed{1}\ \boxed{1}\ \boxed{1}\ \boxed{1}$

0 - Hilary voters

1 - Bernie voters.

<u>Sum</u> of draws corresponds to # people who voted for Bernie.

↑ these 2 examples are when we're <u>counting</u> outcomes.

<u>Dice.</u>

$\boxed{1}\ \boxed{2}\ \boxed{3}\ \boxed{4}\ \boxed{5}\ \boxed{6}$ ← numbers on a die.

<u>SUM</u> of draws is the total pip count on N rolls of a die.

<u>Making a Box Model.</u>

- which numbers go on the tickets.
- how many tickets of each kind.
- how many draws

Example.

Roulette wheel. 38 numbered slots

1 - 36 half are coloured red
half black.

0
00

Simplest bet is Red or Black.

If Bet Red, and ball lands on red, get stake back plus the same amount

ie if bet \$1 on red, and win, get \$2 back

On each spin, gain is either -1
$+1$

These are the numbers that go on the tickets.

18 Red slots $\Rightarrow$ 18 tickets with $+1$

20 other slots (18 black, 0, 00) $\Rightarrow$ 20 tickets with -1.

18 tickets $\boxed{+1}$ 20 tickets with $\boxed{-1}$

ie you lose with prob $\frac{20}{38} > \frac{1}{2}$.

Each spin is now one draw with replacement from this box.

\# of draws from the box $\equiv$ # times we play.

Expected Value.

draw repeatedly with replacement from the box

	R	R	R	B	B	R	R	G	B	R.
win / lose	+1	+1	+1	−1	−1	+1	+1	−1	−1	+1
net gain	1	2	3	2	1	2	3	2	1	2.

Repeat the 10 spins a large number of times.

- the net gain will fluctuate.
  - quantify the value about which it fluctuates — expected value.
  - quantify the size of the fluctuations — standard error. (SE).

- get these quantities from the box model.

Imagine playing roulette 38 times.

should expect 18 $\boxed{+1}$ and 20 $\boxed{-1}$

expected value is $18 - 20 = -2$.

<u>in general</u>

Expected value for the <u>sum</u> of draws made at random with replacement

is ~~# draws~~

(# draws) × (average of box)

Roulette, 38 spins

$$38 \times \left(\frac{18-20}{38}\right) = -2.$$

100 tosses of a coin

$\boxed{0}$ $\boxed{1}$ sum of draws = # H

T H

expected value of # H in 100 tosses is

$$100 \times \left(\frac{0+1}{2}\right). = 50.$$

Quantifying the Variability

Most times that we toss a coin 100 times we will not get exactly 50 H.

$$\begin{array}{c}\# \text{ H in} \\ \text{100 tosses}\end{array} = \begin{array}{c}\text{expected} \\ \text{value} \\ (50)\end{array} + \begin{array}{c}\text{chance} \\ \text{variation}\end{array}$$

size of chance variation is S.E.

SE for the sum of draws with replacement from a box model is

$$SE = \sqrt{\#\text{draws}} \times (\text{SD of box})$$

Eg 100 tosses of a coin

$$\boxed{\boxed{0}\quad\boxed{1}}$$
T H

sum of draws $\equiv$ # H.

mean of box $\frac{0+1}{2} = \frac{1}{2}.$

SD of box $\sqrt{\frac{(0-\frac{1}{2})^2 + (1-\frac{1}{2})^2}{2}} = \frac{1}{2}.$

100 tosses:

$$\text{expected value} = 100 \times \frac{1}{2} = 50$$

$$SE = \sqrt{100} \times \frac{1}{2} = 5$$

In 100 tosses, we expect 50 ± 5 heads.

<u>Roulette.</u>

$$\boxed{18 \text{ tickets } \boxed{+1} \qquad 20 \text{ tickets } \boxed{-1}}$$

$$\text{mean of box} = \frac{18 \times 1 + 20 \times -1}{38} = \frac{-1}{19}$$

SD of box — SD of list of 38 numbers, 18 of which are +1
20 of which are −1

$$\sqrt{\left[\frac{\left[\left(1-\left(\frac{-1}{19}\right)\right)^2 + \left(1-\left(\frac{-1}{19}\right)\right)^2 + \ldots \text{(18 terms)}. + \left(-1-\left(\frac{-1}{19}\right)\right)^2 + \left(-1-\left(\frac{-1}{19}\right)\right)^2 + \ldots \text{20 terms}\right]}{38.}\right]}$$

$= \underline{0.9986}.$ ← fluctuations will be <u>large</u>.

10 spins

expected value $10 \times \frac{-1}{19} = -\frac{10}{19} = -0.5263$

standard error $\sqrt{10} \times 0.9986 = 3.16$

fluctuations are very large

- you can be ahead for a long time.

Note.

observed values are rarely more than 2-3 SE from the expected value.

If everyone in class tossed a coin 100 times

What % of you will have # heads > 55?

- use the Normal approximation.
- convert 55 to standard units.

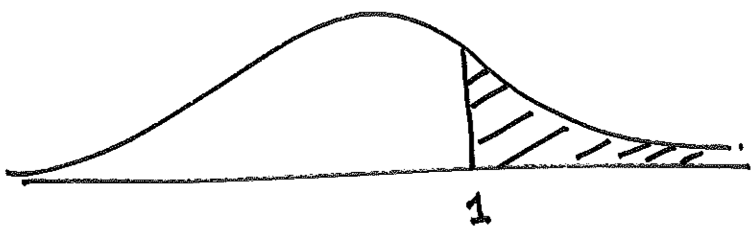
subtract the expected value

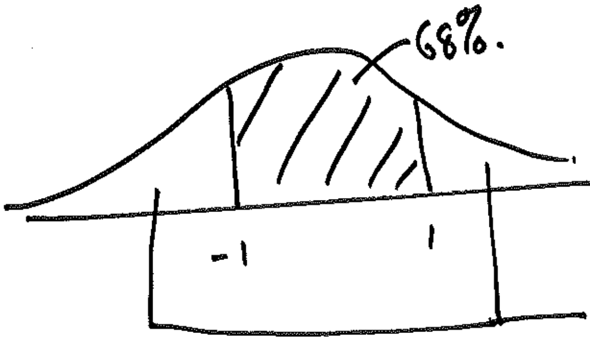
divide by the standard error.

convert 55 to standard units.

$$\frac{55-50}{5} = 1.$$

$$\frac{\text{observed value} - \text{expected value}}{SE}$$





together these are $100 - 68 = 32\%$.

one tail is $\frac{32}{2} = 16\%$.

Expect 16% of you to have ~~or~~ more than 55 H when you toss a coin 100 times.

Rolling Dice.

Roll die 10 times

What's the chance of total # pips < 25 ?

$$\boxed{1}\ \boxed{2}\ \boxed{3}\ \boxed{4}\ \boxed{5}\ \boxed{6}$$

Mean of box ~~expected value~~. $\frac{1+2+3+4+5+6}{6} = \frac{21}{6} = 3.5$

$$SD_{box} \quad \sqrt{\frac{(1-3.5)^2 + (2-3.5)^2 + (3-3.5)^2 + (4-3.5)^2 + (5-3.5)^2 + (6-3.5)^2}{6}}$$

$$= 1.71$$

Expected value in 10 rolls $= 10 \times 3.5 = 35$

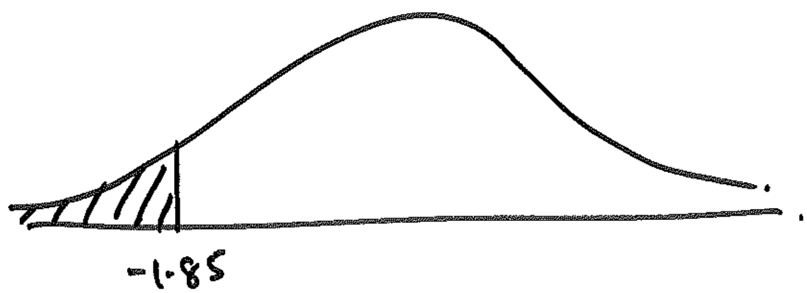
Standard error in 10 rolls $= \sqrt{10} \times 1.71 \approx 5.4$

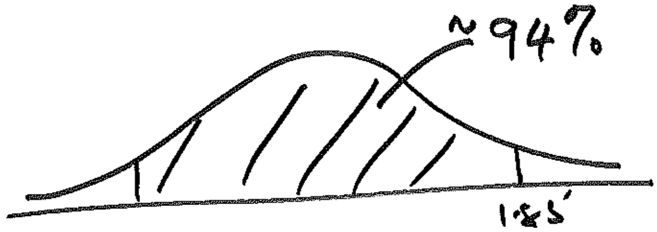
(this looks visually to correspond with the plot of the dice rolls we did in the 1st lecture).

chance of total < 25 ?

convert to standard units

$$\frac{25-35}{5.4} = \frac{-10}{5.4} = -1.85$$





Two tails together are $100 - 94 = 6\%$

Each tail is 3%.

chance of < 25 on 10 rolls of a die

$\simeq 3\%$.

~~later~~ later - when is it reasonable to use the Normal Approximation

Recall Roulette Box

$\boxed{18 \text{ tickets } \boxed{+1} \qquad 20 \text{ tickets } \boxed{-1}}$

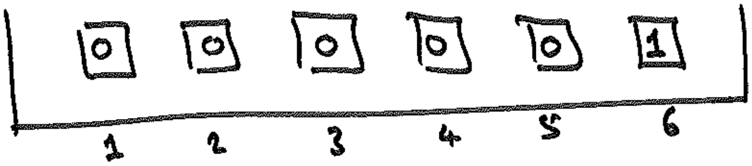
If there are only two different types of tickets

$$SD = \begin{pmatrix} \text{big} & - & \text{small} \\ \text{number} & & \text{number} \end{pmatrix} \times \sqrt{\begin{matrix} \text{fraction} \\ \text{with} \\ \text{big number} \end{matrix} \times \begin{matrix} \text{fraction} \\ \text{with} \\ \text{small number} \end{matrix}}$$

$$= \left(1 - (-1)\right) \times \sqrt{\frac{18}{38} \times \frac{20}{38}}.$$

$$= 0.9986.$$

In 100 rolls of a die, how many 6's should there be?



Put numbers on the tickets so that the <u>SUM</u> of the draws is the number of 6's.

expected value $= 100 \times \underset{\text{of box}}{\text{mean}} = 100 \times \frac{1}{6} = 16\frac{2}{3}$

standard error $= \sqrt{100} \times \text{SD box}$

$= \sqrt{100} \times (1 - 0)\sqrt{\frac{1}{6} \times \frac{5}{6}}$

$= 3.7$

Expect $16\frac{2}{3} \pm 3.7$ sixes in 100 rolls of a die.