Example 1. 250 tickets are drawn at random with replacement from the box

$$\boxed{\boxed{1}\ \boxed{1}\ \boxed{2}\ \boxed{3}\ \boxed{3}}$$

Approximately what will the *sum* of the draws be?

(a) In 250 draws we expect to see:

- $\boxed{1}$ about 100 times (40% of the draws),
- $\boxed{2}$ about 50 times (20% of the draws) and
- $\boxed{3}$ about 100 times.

$\Rightarrow$ Sum of 250 draws $\approx 100 \cdot 1 + 50 \cdot 2 + 100 \cdot 3 = 500$.

Or...

(b) The average of the box is $\dfrac{1+1+2+3+3}{5} = 2$.

$\Rightarrow$ Each draw adds about 2 to the sum (on average).

$\Rightarrow$ Sum of 250 draws $\approx 250 \cdot 2 = 500$.

Definition: The ***Expected Value*** for the sum of n draws ($ARWR$) from a box of numbered tickets is

$$EV(\text{sum}) = n \cdot \text{Avg(box)}.$$

Intuition: Each draw from the box adds *about* Avg(box) to the sum, so

$$\text{sum} \approx \overbrace{\text{Avg(box)} + \text{Avg(box)} + \cdots + \text{Avg(box)}}^{n} = n \cdot \text{Avg(box)}$$

This generalizes explanation **(b)** from Example 1.

Comment: 'Expected value' does not necessarily mean likely value or even possible value...

Another interpretation:

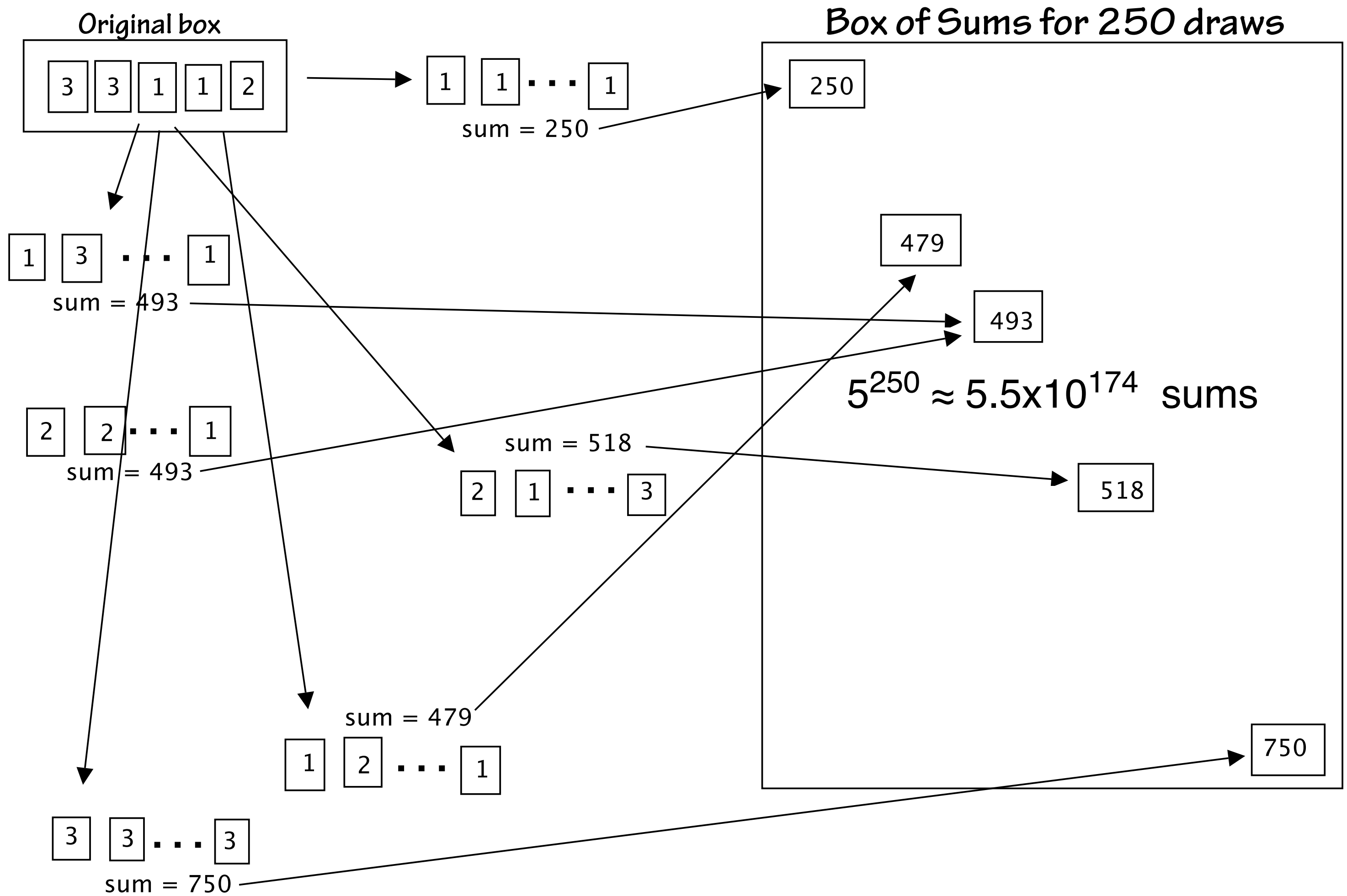
- The box contains m numbered tickets.
- List all the possible combinations of n tickets, and for each one find the sum of the tickets.
- Write each sum on a (new) ticket, and put each sum-ticket into a (new) box — the ***Box of Sums***.

$\Rightarrow$ There are m^n possible ways to draw n tickets *arwr*.

$\Rightarrow$ The Box of Sums contains m^n tickets.

- Drawing n tickets $ARWR$ from the original box and observing the sum is the same as drawing 1 ticket from the box of sums and observing the number.
- *The average of the Box of Sums is equal to n·Avg(original box) because … math.*

Conclusion: ***The Expected Value of the sum of n draws is the average of all possible sums of n draws.***



The *Box of Sums* for Example 1 contains $5^{250} \approx 5.5 \times 10^{174}$ tickets, with numbers ranging from 250 (250 $\boxed{1}$s) to 750 (250 $\boxed{3}$s), and everything in between.

Point of reference: The number of atoms in the universe is $\approx 10^{80}$.

Example 2. In the game *Toss-3*, the player tosses a coin 3 times and receives a payoff based on the number of H are observed:

- 0 H — \$0.
- 1 H — \$1.
- 2 H — \$2.
- 3 H — \$5.

The game costs \$2 to play. If Joe plays *Toss-3* 100 times, how much can he expect to win or lose?

To answer, we create a box model:

- There are 8 possible outcomes when tossing a coin 3 times:

$$TTT \quad TTH \quad THT \quad HTT \quad HHT \quad HTH \quad THH \quad HHH$$

- For each outcome there is a numbered ticket—the number being the profit/loss associated with that outcome:

$$\boxed{-2} \quad \boxed{-1} \quad \boxed{-1} \quad \boxed{-1} \quad \boxed{0} \quad \boxed{0} \quad \boxed{0} \quad \boxed{3}$$

- Playing *Toss-3* is like drawing one ticket from a box containing these tickets.
- Playing 100 times is like drawing from this box 100 times, *ARWR*.
- The expected winnings/losses is equal to the expected value of the sum of these 100 draws.

$$\Rightarrow \text{Avg(box)} = \frac{(-2)+(-1)+(-1)+(-1)+0+0+0+3}{8} = -\frac{1}{4}.$$

$$\Rightarrow \text{EV(sum)} = 100 \times (-0.25) = -25.$$

Answer: Joe can expect to *lose* about \$25 in 100 games of *Toss-3*.

Question: Can we add a *give-or-take* amount to that answer?

(*) n tickets are drawn *ARWR* from a box of numbered tickets with Avg(box)$= A$.

- Each draw adds *about* A to the sum,

$$\Rightarrow \text{sum} \approx n \cdot A = EV(\text{sum}).$$

- "*about A...*" means "$A\pm$(chance error)"

- The chance error for each draw is equal to the SD of the (original) box, on average.

- So: each draw adds about $A \pm SD$ to the sum.

$$\Rightarrow \text{sum} \approx n \cdot (A \pm SD) = EV(sum) \cancel{\pm n \cdot SD} \pm \overbrace{\sqrt{n} \cdot SD}^{\text{chance error}\checkmark}$$

- **Problem:** "$n \cdot SD$" ignores *cancellation* — some positive chance errors cancelling some negative chance errors in the sum.

- **Correction:** The size of the chance error for the sum is $\sqrt{n} \cdot SD$.

Definition: The ***Standard Error*** (SE) for the sum of n tickets, drawn *ARWR* from a box of numbered tickets is

$$SE(\text{sum}) = \sqrt{n} \cdot SD(\text{box})$$

(*) The SE for the sum of n draws is ***the SD of the box of sums.***

Example 2. (continued) The standard deviation for the *Toss-3* box is

$$SD = \sqrt{\frac{(-2-(-0.25))^2+(-1-(-0.25))^2+\cdots+(0-(-0.25))^2+(3-(-0.25))^2}{8}}$$

$$= \sqrt{\frac{31}{16}} \approx 1.392$$

$\Rightarrow$ The standard error for the sum of 100 draws from the *Toss-3* box is

$$SE = \sqrt{100} \cdot \sqrt{31/16} \approx 13.92.$$

$\Rightarrow$ In 100 games of *Toss-3*, Joe can expect to ***lose*** between

$$\$25 - \$13.92 = \$11.08 \quad \text{and} \quad \$25 + \$13.92 = \$38.92.$$

We can restate the last result (slightly) to say:

In 100 games, Joe is ***likely*** lose between \$11.08 and \$38.92.

Question: How likely?

Fun(damental) fact:

> *When the number n of draws is large enough, the tickets in the* Box-of-Sums *have an approximately normal distribution with $Average = EV(sum)$ and $SD = SE(sum)$.*

$\Rightarrow$ When the number of draws is large enough, we can answer probabilistic questions about the sum of draws by converting to standard units and using the normal curve.

Example 2. (continued some more)

- The range 25 ± 13.92 is EV(sum) $\pm$ 1 SE(sum). Assuming that $n = 100$ is large enough, the chance of falling in this range is roughly equal to the area under the normal curve from -1 to 1. I.e., there is roughly a 68% chance that Joe loses between \$11.08 and \$38.92 if he plays *Toss-3* 100 times.

- The probability that Joe doesn't lose any money in 100 games (and perhaps wins some) is

$$P(\text{sum} \geq 0) \approx \text{area under normal curve to right of } z_0,$$

where z_0 = standard units for (sum = 0).

$$\Rightarrow \quad z_0 = \frac{0 - (-25)}{13.92} \approx 1.8$$

so the probability of not losing any money in 100 games is

$$\frac{100\% - T(1.8)}{2} = \frac{100\% - 92.81\%}{2} = 3.595\%$$

Counting and Sums: zero-one boxes.

Observations:

- If the box of tickets contains only $\boxed{1}$s and $\boxed{0}$s, then the ***sum*** of n draws from the box is equal to the ***number*** of $\boxed{1}$s in the draw.

- If N is the number of tickets in the box and p is the proportion of $\boxed{1}$s in the box, then there are Np $\boxed{1}$s and $N(1-p)$ $\boxed{0}$s and the average of the box is

$$\frac{\overbrace{1+1+\cdots+1}^{Np}+\overbrace{0+0+\cdot+0}^{N(1-p)}}{N}=\frac{Np}{N}=p.$$

- The SD of this box is $\sqrt{p(1-p)}$.

Because... ... *of a little more arithmetic.* (See slides from April 9).

Example 3. A fair coin is tossed 200 times. What is the probability that we observe between 90 and 110 H?

- The number of H in 200 tosses of a fair coin can be modeled by the number of $\boxed{1}$s in n draws *ARWR* from a 0-1 box...
- ... which is like the *sum* of 200 draws...
- ... *'Box of sums'* has an ***approximately normal distribution***, with

$$Avg = EV(sum) = \frac{1}{2} \cdot 200 = 100$$

and

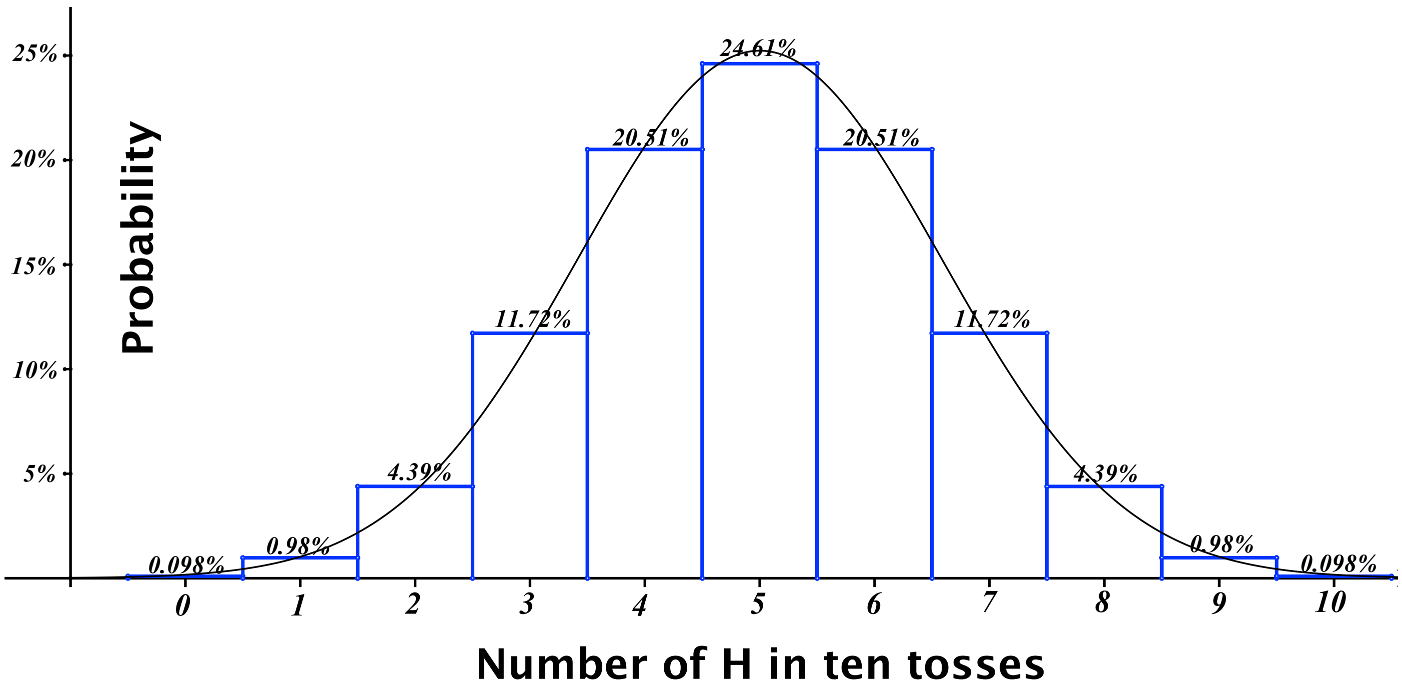
$$SD = SE(sum) = \sqrt{200} \cdot \sqrt{\frac{1}{2} \cdot \frac{1}{2}} \approx 7.07.$$

So, the probability that we observe between 90 and 110 H is roughly equal to the area under the normal curve between

$$\frac{90-100}{7.07} \approx -1.414 \quad \text{and} \quad \frac{110-100}{7.07} \approx 1.414.$$

which is approximately 84%. *Or is it...?*

Remember this picture?



(*) Bars start 0.5 below each number and end 0.5 above.

(*) To answer probability questions about *counting*, like "*what are the chances that the number of* $\boxed{1}$ *s in* n *draws is between* a *and* b", we convert $a - 0.5$ and $b + 0.5$ to standard units and proceed as before.

(*) Assumptions: a, b are both integers and $a \leq b$.

Example 3. (corrected)

... the probability that we observe between 90 *and* 110 H *is roughly equal to the area under the normal curve between*

$$\frac{89.5 - 100}{7.07} \approx -1.485 \quad \text{and} \quad \frac{110.5 - 100}{7.07} \approx 1.485.$$

which is about **86%**.