(*) If 4 tickets are drawn with replacement from | 1 | 2 | 2 | 4 | 6 |, what are the chances that we observe *exactly* two [2]s?

$\Rightarrow$ *'Exactly two'* [2]s in a sequence of four draws can occur in many ways. For example, ([2] - [*not* 2] - [*not* 2] - [2]), ([2] - [2] - [*not* 2] - [*not* 2]), ([2] - [*not* 2] - [2] - [*not* 2]), and so on.

Two key observations:

(i) All these different sequences are mutually exclusive of each other. This is because, if we observe the sequence ([2] - [*not* 2] - [2] - [*not* 2]), for example, then we do not observe the sequence ([2] - [*not* 2] - [*not* 2] - [2]).

(ii) The probability of observing each of these sequences is *the same* for all of them, because

$$\frac{2}{5} \cdot \frac{3}{5} \cdot \frac{3}{5} \cdot \frac{2}{5} = \frac{2}{5} \cdot \frac{2}{5} \cdot \frac{3}{5} \cdot \frac{3}{5} = \frac{2}{5} \cdot \frac{3}{5} \cdot \frac{2}{5} \cdot \frac{3}{5} = \cdots = 5.76\%$$

This means that

$$P(\text{exactly two } \boxed{2}\text{s in four draws}) = \overbrace{\frac{36}{625} + \frac{36}{625} + \frac{36}{625} + \cdots + \frac{36}{625}}^{\text{number of sequences with two } \boxed{2}\text{s}}$$

The only thing that remains is to figure out how many sequences there are with exactly two $\boxed{2}$s...

Observations.

(i) We don't care which tickets go in the 'not $\boxed{2}$' spots.

(ii) Since we are (theoretically) listing **all** of the possible 2-$\boxed{2}$ sequences, we don't need to think about this process as a bunch of 'random draws'... we can be methodical.

(iii) When listing different 2-$\boxed{2}$ sequences, all we have to decide is where in each sequence to put the $\boxed{2}$s... the 'not $\boxed{2}$'s will go in the other two spots.

$\Rightarrow$ *The number of different sequences with two* $\boxed{2}$ *s is equal to the number of ways to choose two positions in a sequence of four.*

$\Rightarrow$ There are 4 positions in which we can place the first $\boxed{2}$, and for each choice of first position, there are 3 ways to choose the second position...

So it seems that there are $4 \cdot 3 = 12$ ways to place two $\boxed{2}$s in a sequence of four draws...

But we are *overcounting*, because each pair of positions has been counted *twice*! For example, the choices *'first* $\boxed{2}$ *in the third position and second* $\boxed{2}$ *in the first position'* and *'first* $\boxed{2}$ *in the first position and second* $\boxed{2}$ *in the third position'* result in the *same* pair of positions — first and third.

Conclusion: The number of sequences with exactly two $\boxed{2}$s is $\dfrac{4 \cdot 3}{2} = 6$... So

$$P(\text{exactly two } \boxed{2}\text{s in four draws}) = \overbrace{\frac{36}{625} + \frac{36}{625} + \frac{36}{625} + \cdots + \frac{36}{625}}^{\text{number of sequences with two } \boxed{2}\text{s}}$$

$$= \left(\frac{36}{625}\right) \times 6 = \frac{216}{625} = 34.56\%$$

More general question: If n tickets are drawn at random with replacement from the box

$$\boxed{1}\ \boxed{2}\ \boxed{2}\ \boxed{4}\ \boxed{6},$$

what are the chances that exactly k of them will be $\boxed{2}$s?

The reasoning that we used when $n = 4$ and $k = 2$ can be used to answer this question too.

(*) The results of different draws are *independent*.

(*) The probability of a $\boxed{2}$ on any one draw is 2/5.

(*) The probability of a *not* $\boxed{2}$ on any one draw is 3/5.

(*) I will henceforth label '*not* $\boxed{2}$' by $\boxed{?}$.

Intermediate conclusion 1.

The probability of any particular sequence of n draws which results in k $\boxed{2}$s and $(n-k)$ $\boxed{?}$s

$$\overbrace{\boxed{?}\ \boxed{?}\ \boxed{2}\ \boxed{?}\ \boxed{2}\ \cdots\ \boxed{?}\ \boxed{2}\ \boxed{?}}^{k\ \boxed{2}\ s \text{ and } (n-k)\ \boxed{?}\ s}$$

is equal to

$$\overbrace{\frac{3}{5}\cdot\frac{3}{5}\cdot\frac{2}{5}\cdot\frac{3}{5}\cdot\frac{2}{5}\cdots\frac{3}{5}\cdot\frac{2}{5}\cdot\frac{3}{5}}^{k\ 2/5\,s \text{ and } (n-k)\,3/5\,s} = \left(\frac{2}{5}\right)^k \cdot \left(\frac{3}{5}\right)^{n-k}$$

regardless of the order in which the tickets appear!

(*) Different sequences of k $\boxed{2}$s and $(n-k)$ $\boxed{?}$s (i.e., sequences that differ in at least one position (actually, at least two)) are *mutually exclusive*.

(*) We can use the addition rule to conclude that

$$P(\text{exactly } k \; \boxed{2} \; s \text{ in } n \text{ draws})$$

$$= \overbrace{\left(\frac{2}{5}\right)^k \cdot \left(\frac{3}{5}\right)^{n-k} + \left(\frac{2}{5}\right)^k \cdot \left(\frac{3}{5}\right)^{n-k} + \cdots + \left(\frac{2}{5}\right)^k \cdot \left(\frac{3}{5}\right)^{n-k}}^{\#\text{ of different sequences with exactly } k \; \boxed{2} \text{s and } (n-k) \; \boxed{?} \text{s}}$$

$$= (\text{Unknown number}) \cdot \left(\frac{2}{5}\right)^k \cdot \left(\frac{3}{5}\right)^{n-k}.$$

Next Question: *What is the 'unknown number'?*

I.e., how many sequences of draws are there with k $\boxed{2}$s and $(n-k)$ $\boxed{?}$s?

(*) We only need to count the number of ways of choosing k positions for the $\boxed{2}$s among the n available positions.

$$\overbrace{\boxed{?}\,\boxed{?}\,\boxed{2}\,\boxed{?}\,\boxed{2}\,\cdots\,\boxed{?}\,\boxed{2}\,\boxed{?}}^{n \text{ positions and } k\ \boxed{2}\ s}$$

- There are $n \cdot (n-1) \cdot (n-2) \cdots (n-k+1)$ different ways that we can place the $\boxed{2}$s *if the order matters*: first $\boxed{2}$, second $\boxed{2}$, etc.
- But we don't care about the order in which the positions were chosen, so the number above is too big — we are counting each of the possible sequences too many times.
- Every ***unordered set*** of k positions of the $\boxed{2}$s appears

$$k! = k \cdot (k-1) \cdots 2 \cdot 1$$

different times in the collection of *ordered* sets we counted above.

Intermediate conclusion 2.

The number of sequences of n draws that result in k $\boxed{2}$s and $(n-k)$ $\boxed{?}$s is equal to

$$\frac{n\cdot(n-1)\cdot(n-2)\cdots(n-k+1)}{k!} = \frac{n!}{(n-k)!\cdot k!} = \binom{n}{k}.$$

Final conclusion.

If n tickets are drawn at random with replacement from the box

$$\boxed{\boxed{1}\ \boxed{2}\ \boxed{2}\ \boxed{4}\ \boxed{6}},$$

the probability of observing exactly k $\boxed{2}$s is

$$P(\text{exactly } k\ \boxed{2}\, s \text{ in } n \text{ draws}) = \binom{n}{k}\cdot\left(\frac{2}{5}\right)^{k}\cdot\left(\frac{3}{5}\right)^{n-k}.$$

Comments:

- $\binom{n}{k}$ is pronounced '***n choose k***'. It is the number of different (unordered) subsets of size k that can be chosen from a set of n objects.
- $\binom{n}{0} = 1$ by definition.
- $\binom{n}{k} = \binom{n}{n-k}$.
- The binomial coefficients large quickly. For example,

$$\binom{10}{3} = 120,\ \binom{10}{5} = 252,\ \binom{20}{3} = 1140,\ \binom{20}{5} = 15504$$

  and

$$\binom{100}{30} = 29372339821610944823963760$$

- The numbers $\binom{n}{k}$ are called ***binomial coefficients*** because they appear in the ***binomial formula***

$$(a+b)^n = \binom{n}{0}a^n + \binom{n}{1}a^{n-1}b + \cdots + \binom{n}{k}a^{n-k}b^k + \cdots + \binom{n}{n}b^n.$$

The general case.

Suppose a box contains N tickets, some if which are $\boxed{1}$s and that the probability of (randomly) drawing a $\boxed{1}$ from the box is $P(\boxed{1}) = p$.

$\Rightarrow$ The number of $\boxed{1}$s in the box is $p \cdot N$.

$\Rightarrow$ The probability of drawing a not-$\boxed{1}$ is $1 - p$.

If n tickets are drawn at random with replacement from the box, then the probability of observing exactly k $\boxed{1}$s is

$$P(\text{exactly } k \ \boxed{1}\text{s in } n \text{ draws}) = \binom{n}{k} p^k (1-p)^{n-k}.$$

Observation. The *number* N of tickets in the box is less important here than the *proportion* p of $\boxed{1}$s in the box.

Coin tosses.

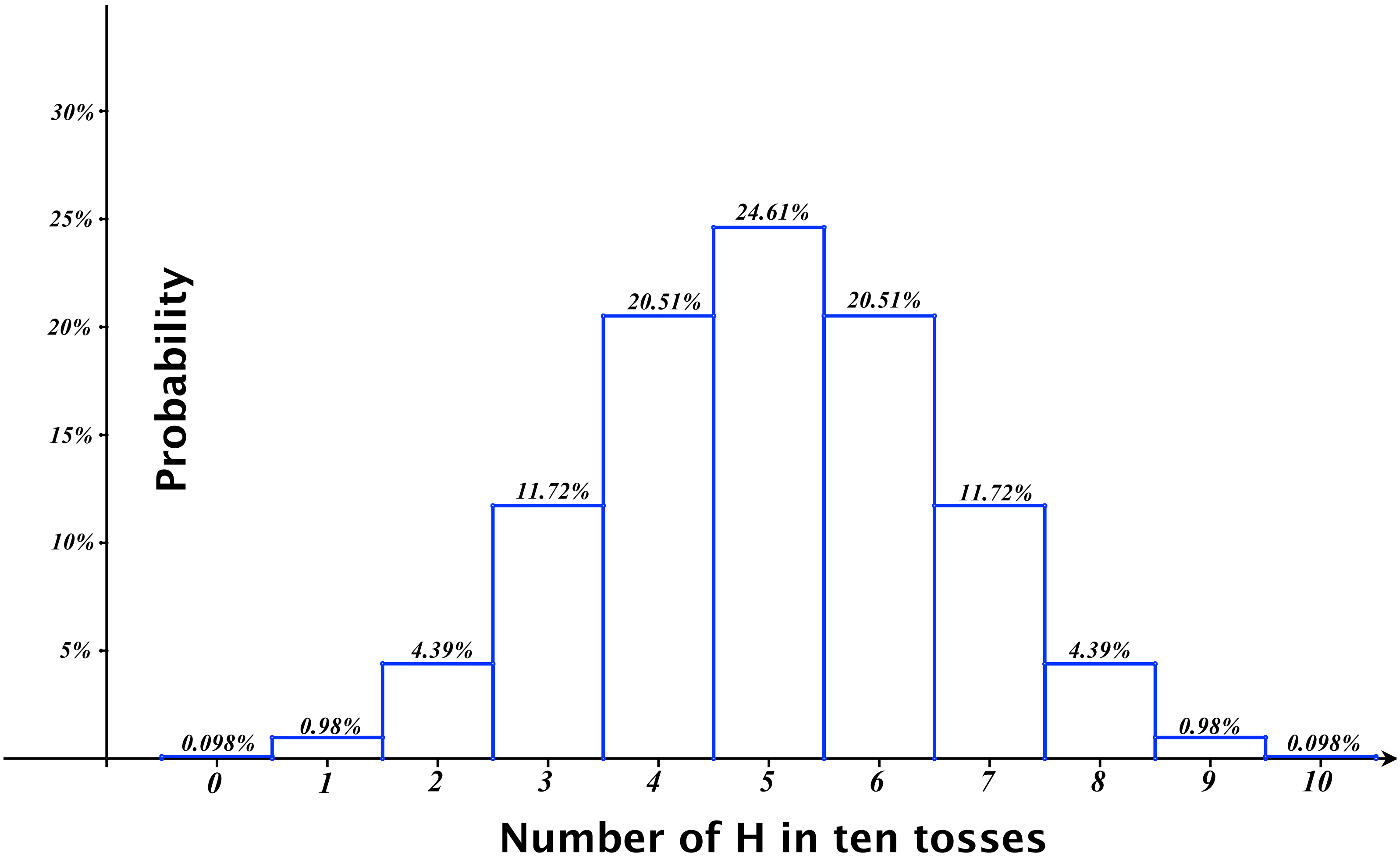
If we have a box with two tickets, for example one $\boxed{1}$ and one $\boxed{0}$, then the number of $\boxed{1}$s in n random draws with replacement from this box can be used to model the number of *heads* in n tosses of a fair coin.

(*) The probability of observing k heads in n tosses of a fair coin is

$$P(k \text{ heads in } n \text{ tosses}) = \binom{n}{k} \cdot \left(\frac{1}{2}\right)^{k} \cdot \left(\frac{1}{2}\right)^{n-k} = \binom{n}{k} \cdot \left(\frac{1}{2}\right)^{n}.$$

(*) Given a particular n, there are $n+1$ possible values for k (i.e., $0, 1, 2, \ldots, n$) and the probabilities for the different values of k can be displayed in a ***probability histogram***.

⇒ The values of k are arranged on the horizontal axis and we use the **density scale** on the vertical axis: ***the area of the bar above each value k gives the probability of observing exactly k heads in n tosses.***



Probability histogram for the number of heads in 10 *tosses of a fair coin.*

We can 'read' this histogram the same way that we do a histogram for data...

(*) What is the probability of observing more than 7 heads in 10 tosses?

⇒ More than 7 heads in 10 tosses means 8 heads, 9 heads or 10 heads, and these are all mutually exclusive events. So...

$$\begin{aligned} &P(\text{more than 7 heads in 10 tosses}) \\ &\quad = P(\text{8 heads}) + P(\text{9 heads}) + P(\text{10 heads}) \\ &\quad = \text{area under histogram from 7.5 to 10.5} \\ &\quad \approx 0.0439 + 0.0098 + 0.00098 \approx 0.0547 \end{aligned}$$

(*) What is the probability of observing between 4 and 6 heads in 10 tosses?

$$\begin{aligned} \Rightarrow\ &P(\text{between 4 and 6 heads in 10 tosses}) \\ &\quad = P(\text{4 heads}) + P(\text{5 heads}) + P(\text{6 heads}) \\ &\quad = \text{area under histogram from 3.5 to 6.5} \\ &\quad \approx 0.2051 + 0.2461 + 0.2051 = 0.6563 \end{aligned}$$

A hint of things to come...