Example. If 4 tickets are drawn with replacement from

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

what are the chances that we observe *exactly* two $\boxed{2}$s?

$\Rightarrow$ *'Exactly two'* $\boxed{2}$s in a sequence of four draws can occur in many ways. For example, ($\boxed{2}$ - $\boxed{*}$ - $\boxed{*}$ - $\boxed{2}$), ($\boxed{2}$ - $\boxed{2}$ - $\boxed{*}$ - $\boxed{*}$), ($\boxed{2}$ - $\boxed{*}$ - $\boxed{2}$ - $\boxed{*}$), and so on. (where $\boxed{*}$ means $\boxed{\textit{not}\ 2}$).

Two key observations:

(i) All these different sequences are mutually exclusive of each other...

If we observe the sequence ($\boxed{2}$ - $\boxed{*}$ - $\boxed{2}$ - $\boxed{*}$), then we do not observe the sequence ($\boxed{2}$ - $\boxed{*}$ - $\boxed{*}$ - $\boxed{2}$).

(ii) The probability of observing each of these individual sequences is *the same* for all of them, because *multiplication is commutative*, i.e.,

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

This means that

$$P(\text{exactly two } \boxed{2} \text{s in four draws}) = \overbrace{5.76\% + 5.76\% + 5.76\% + \cdots + 5.76\%}^{N \,=\, \text{number of sequences with two } \boxed{2}\text{s}}$$
$$= N \times 5.76\%$$

Now we have to figure out what N is...

Observations.

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

(ii) Since we are listing **all** of the possible 2-$\boxed{2}$ sequences, 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 $\Longrightarrow$ the 'not $\boxed{2}$'s will go in the other 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

$$N = \frac{4 \cdot 3}{2} = 6$$

So

$$P(\text{exactly two } \boxed{2}\text{s in four draws}) = 6 \times 5.76\% = 34.56\%.$$

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

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

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 $\boxed{*}$ $\left(\text{not } \boxed{2}\right)$ on any one draw is $3/5$.

Observation 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!

Observation 2.

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*.

This means that we can use the addition rule to conclude that

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

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

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

N = number of different sequences with exactly k $\boxed{2}$s and $(n-k)$ $\boxed{*}$.

Next question: *What is N?*

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.

Observation 3.

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

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

Comment: $\binom{n}{k}$ is one of the standard ways of denoting this number. Another standard notation for this is ${}_nC_k$.

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***', and is also called a *binomial coefficient*. 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 grow 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: $\boxed{1}$s and $\boxed{0}$s. And suppose 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 $\boxed{0}$ is $1 - p$.

If n tickets are drawn at random with replacement from such a box, then the probability of observing exactly k $\boxed{1}$s (and $(n-k)$ $\boxed{0}$s) is

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

Observation. For this question, the *number* N of tickets in the box is not important. What matters is the *proportion* p of $\boxed{1}$s in the box.

Coin tosses.

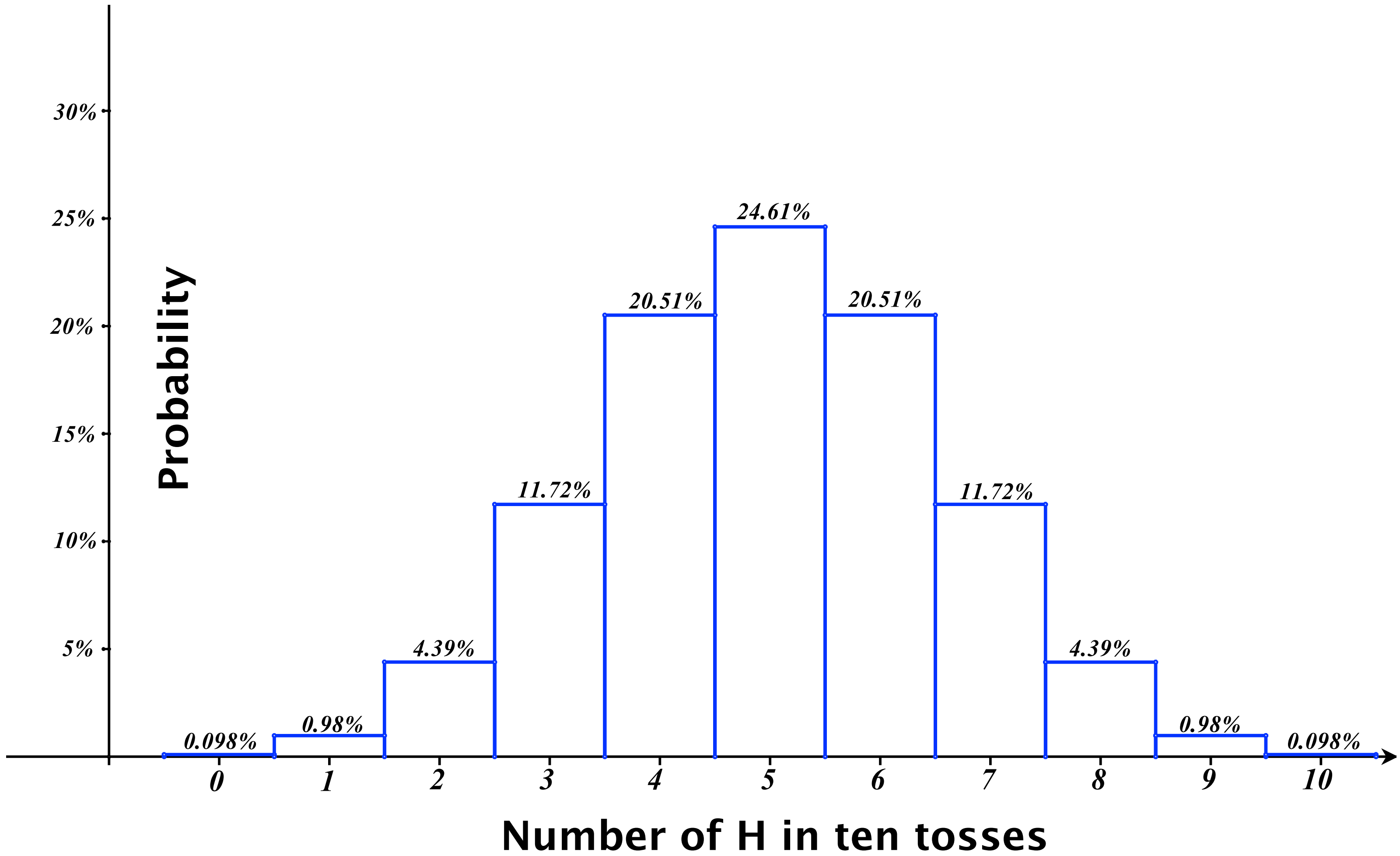
If we have a box with one $\boxed{1}$ and one $\boxed{0}$, then the number of $\boxed{1}$s in n random draws with replacement from this box can be used to model the number of *heads* in n tosses of a fair coin.

(*) 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 each of these values can be displayed in a ***probability histogram***.

$\Rightarrow$ 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?

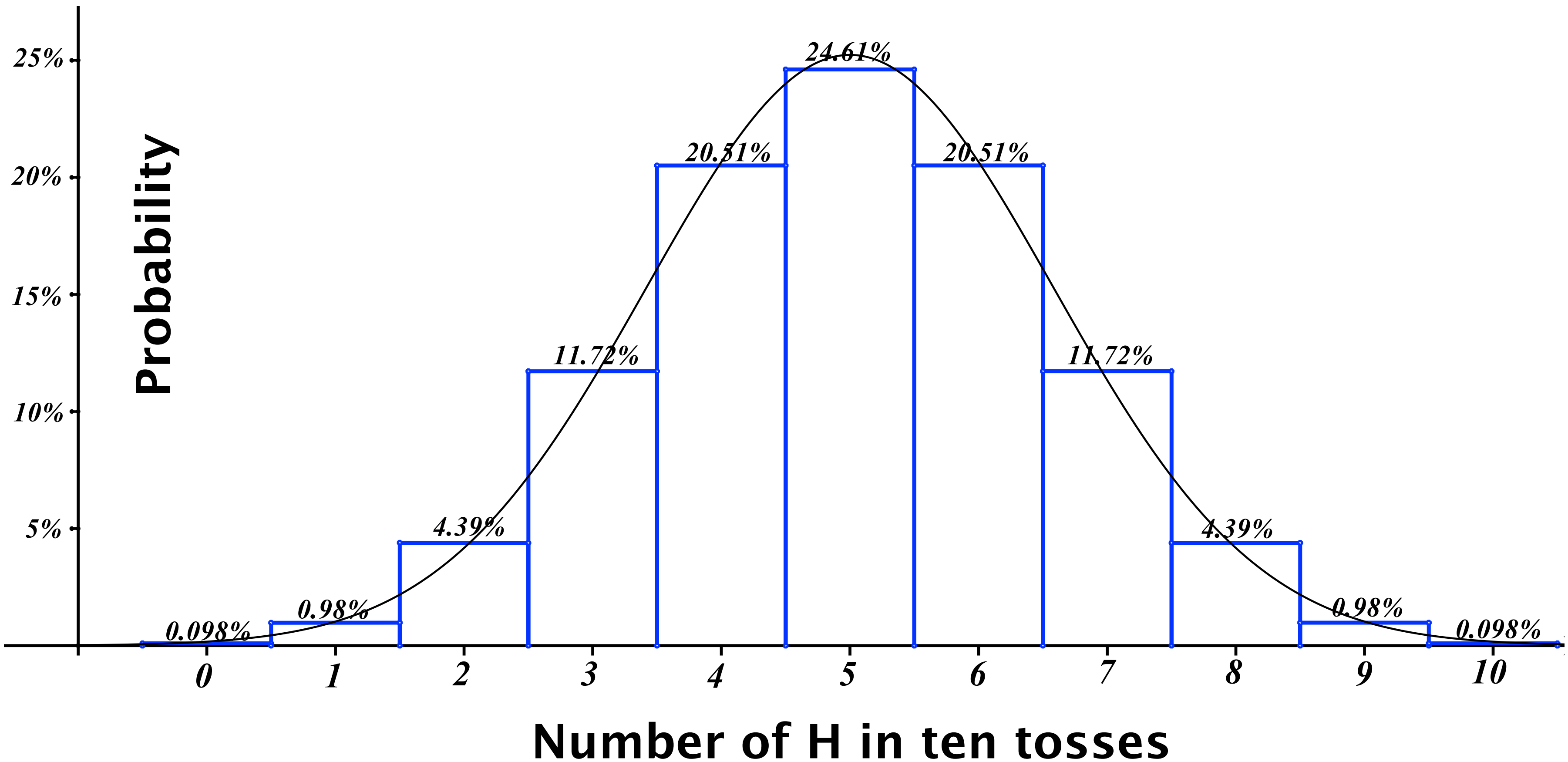
$\Rightarrow$ More than 7 heads in 10 tosses means 8 heads, 9 heads or 10 heads, and these are all mutually exclusive events. So...

$$
\begin{aligned}
&P(\text{more than 7 heads in 10 tosses}) \\
&\quad = P(8 \text{ heads}) + P(9 \text{ heads}) + P(10 \text{ heads}) \\
&\quad = \text{area under histogram from 7.5 to 10.5} \\
&\quad \approx 4.39\% + 0.98\% + 0.098\% \approx 5.47\%
\end{aligned}
$$

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

$$
\begin{aligned}
\Rightarrow\ &P(\text{between 4 and 6 heads in 10 tosses}) \\
&\quad = P(4 \text{ heads}) + P(5 \text{ heads}) + P(6 \text{ heads}) \\
&\quad = \text{area under histogram from 3.5 to 6.5} \\
&\quad \approx 20.51\% + 24.61\% + 20.51\% = 65.63\%
\end{aligned}
$$

A hint of things to come...



... But first, more examples.

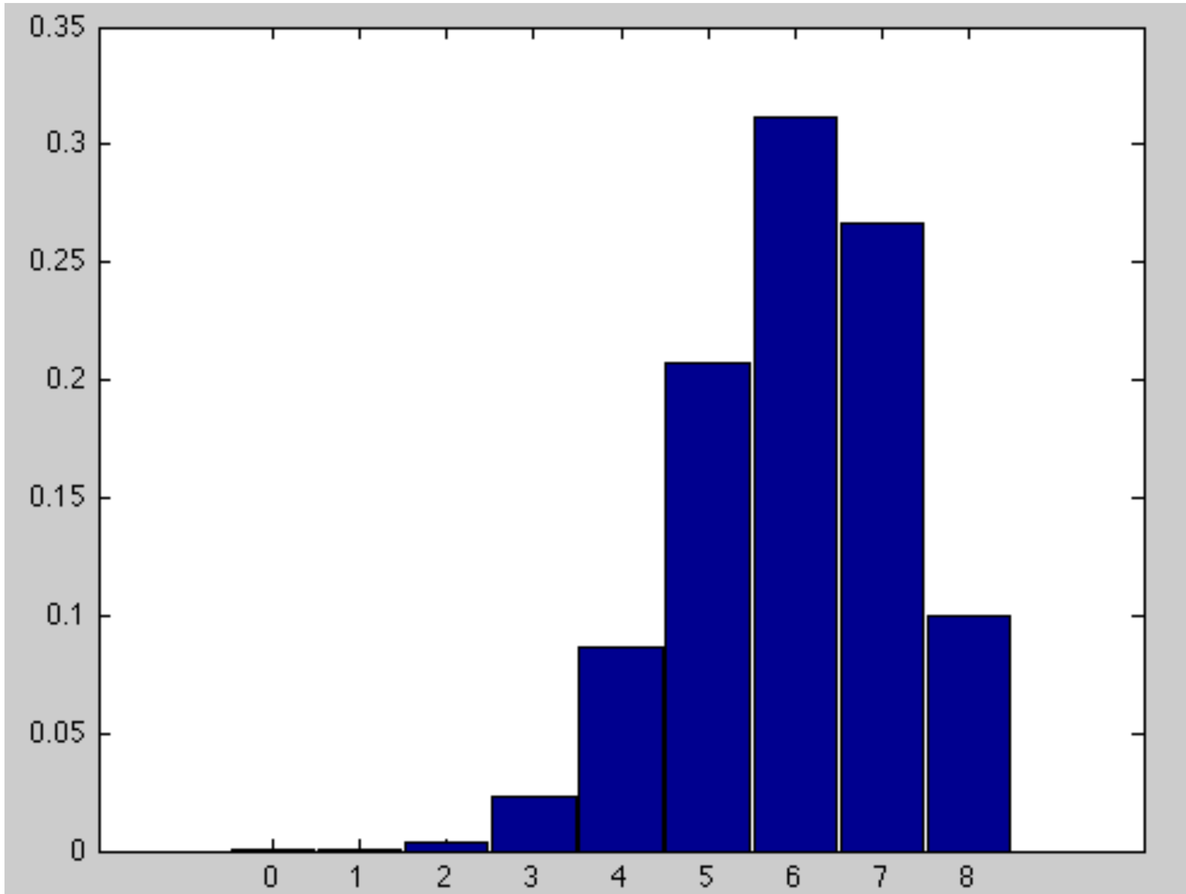
Example. 8 tickets are drawn at random with replacement from a box containing 8 tickets — 6 $\boxed{1}$s and 2 $\boxed{0}$s.

What is the probability that we will observe between 5 and 7 $\boxed{1}$ s?

We can calculate the probability using (the addition rule and) the formula for binomial probabilities:

$$\begin{aligned}
P\left(\text{between 5 and 7 }\boxed{1}\text{s}\right) &= P\left(5\,\boxed{1}\,s\right) + P\left(6\,\boxed{1}\,s\right) + P\left(7\,\boxed{1}\,s\right) \\
&= \binom{8}{5}\cdot\left(\frac{3}{4}\right)^5\cdot\left(\frac{1}{4}\right)^3 + \binom{8}{6}\cdot\left(\frac{3}{4}\right)^6\cdot\left(\frac{1}{4}\right)^2 \\
&\qquad + \binom{8}{7}\cdot\left(\frac{3}{4}\right)^7\cdot\left(\frac{1}{4}\right)^1 \\
&= 56\cdot\frac{3^5}{4^8} + 28\cdot\frac{3^6}{4^8} + 8\cdot\frac{3^7}{4^8} \\
&= \frac{51516}{65536} \approx 0.786 = 78.6\%
\end{aligned}$$

(*) We can use a probability histogram in this case too:



$$P\left(\text{between 5 and 7 } \boxed{1}\text{s}\right) = \text{area of histogram between 4.5 and 7.5}$$
$$\approx 0.21 + 0.31 + 0.27 = 0.79$$

Question: *What changes when the number of draws increases?*

(*) 80 tickets are drawn at random with replacement from the same box.

⇒ What is the probability that we will observe between 55 and 65 $\boxed{1}$ s?

⇒ We can give a precise answer using the binomial formula:

$$P\left(\text{ between 55 and 65 } \boxed{1} \text{s in 80 draws}\right)$$

$$= \binom{80}{55} \cdot \left(\frac{3}{4}\right)^{55} \cdot \left(\frac{1}{4}\right)^{25} + \binom{80}{56} \cdot \left(\frac{3}{4}\right)^{56} \cdot \left(\frac{1}{4}\right)^{24} + \cdots$$

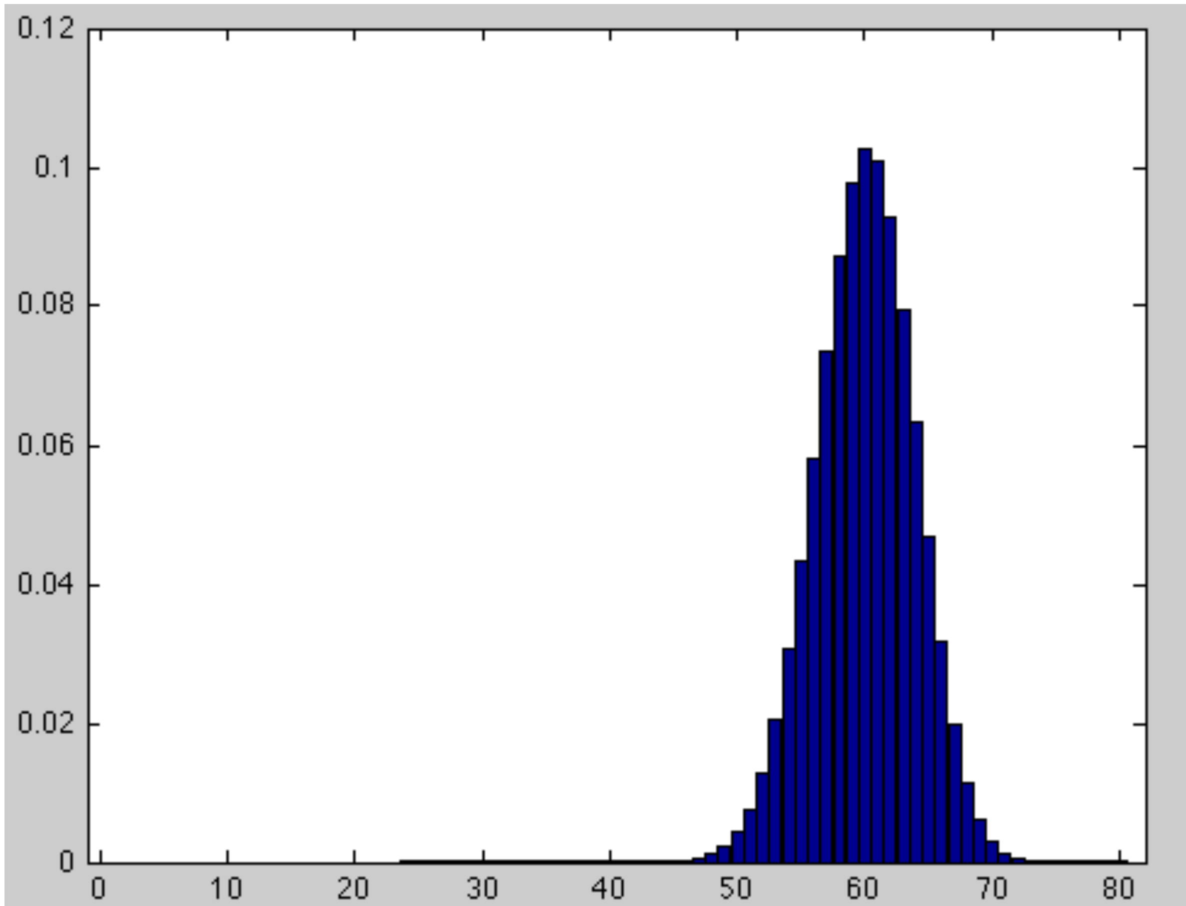
$$\cdots + \binom{80}{65} \cdot \left(\frac{3}{4}\right)^{65} \cdot \left(\frac{1}{4}\right)^{15} = \ldots ?$$

These days, evaluating expressions like this directly is easy with computers. For example, we can use on-line calculators like the one found here:

http://stattrek.com/online-calculator/binomial.aspx

Answer: $\approx 84.6\%$.

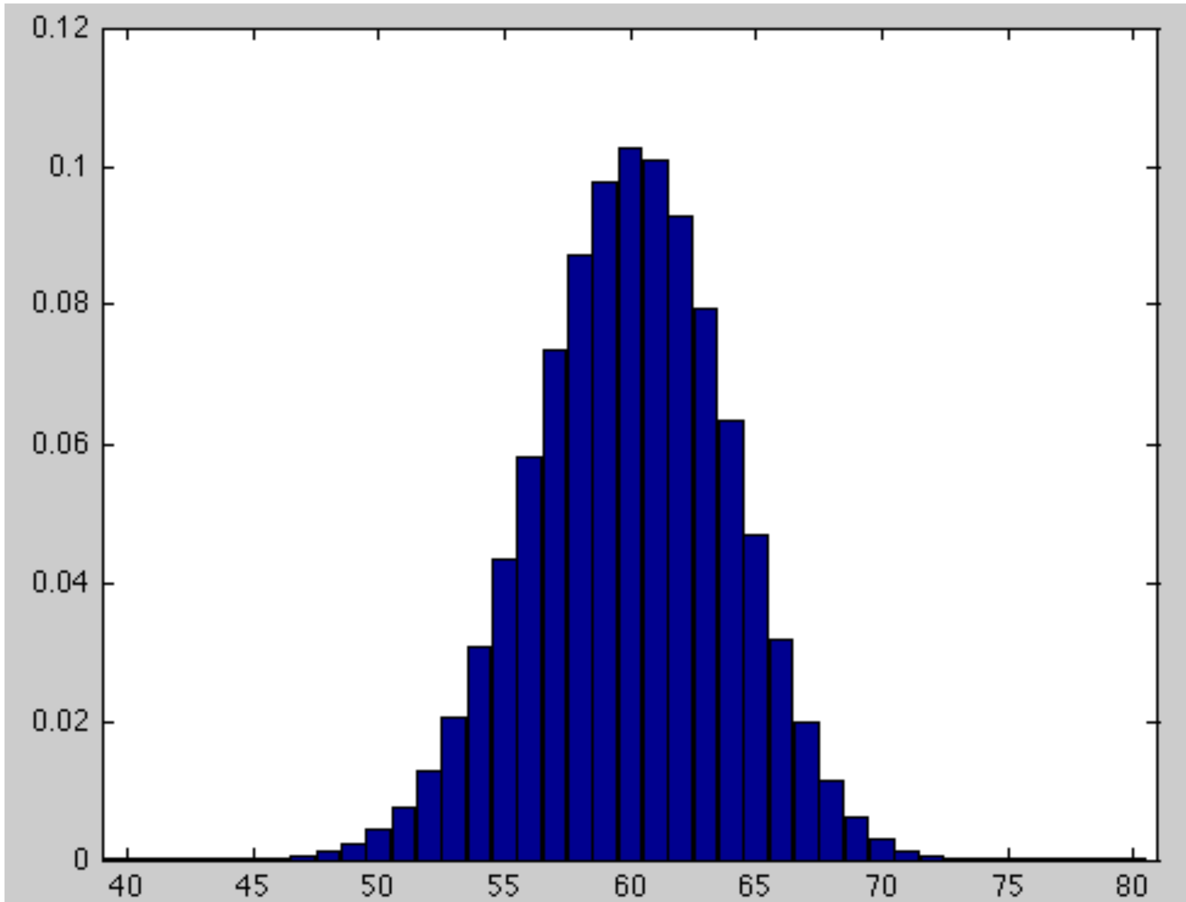
What about using a probability histogram for the number of [1]s observed in 80 random draws (with replacement) from the same box..?



(*) Kind of hard to read.

(*) Also - we have to calculate all the probabilities to draw the histogram!

'Zooming in' to the part of the histogram where 40 $\leq$ number of $\boxed{1}$s $\leq$ 80:



(*) This part of the histogram accounts for more than 99.9999% of the total.

(*) I.e., P(fewer than 40 $\boxed{1}$s in 80 draws) $< 0.0001\%$.

Another hint at things to come...