Example 1. A card is dealt from the top of a well-shuffled deck, then it is replaced, the deck is reshuffled and another card is dealt. What is the probability that the second card is a 7 given that the first card is a King?

$\Rightarrow$ Since the first card was replaced (and the deck was reshuffled) before the second card was dealt, the nature of the first card doesn't provide any new information about the nature of the second card, so

$$P(\text{second card } 7|\text{first card King}) = P(\text{second card } 7) = \frac{4}{52} \approx 7.7\%.$$

Definition. If $P(E|F) = P(E)$, then the events E and F are said to be (statistically) ***independent***.

Comment: *Independence* is not the same as unrelated. Two events can be closely related, but statistically independent.

Example 2. A box contains 200 tickets...

- 120 of the tickets are marked with an X and 80 tickets are marked with a Y.
- Of the X-tickets, 30 are also marked with an A and the other 90 are marked with a U.
- Of the Y-tickets, 20 are also marked with an A and the other 60 are marked with a U.

One ticket is drawn at random from the box...

(*) There are 200 tickets overall and $50 = 30 + 20$ of them are marked with an A, so $P(A) = 25\%$.

(*) There are 120 X-tickets and 30 of them are marked with an A, so $P(A|X) = 30/120 = 25\%$.

(*) The events *"ticket is marked with an A"* and *"ticket is marked with an X"* are *independent* (but not unrelated).

Comments.

(1) If the events E and F are independent, then $P(E|F) = P(E)$ (and $P(F|E) = P(F)$). So, if E and F are independent events, then the multiplication rule reduces to

$$P(E \text{ and } F) = P(E)P(F).$$

In fact, this formula can be used as the definition of independence.

(2) Independence is usually an assumption that we make about the events that we are considering because its makes computing probability easier.

(3) But — ***the assumption of independence must be justified in one way or another***, and reevaluated if subsequent results point in another direction. One of the easiest ways to misuse probability is to make an unjustified assumption of independence. See the example in Section 13.5 of the textbook.

Box models.

Many questions in probability can be answered by considering an appropriate *box model.*

A box model is comprised of two components.

(i) A (hypothetical) box of tickets, each of which is labelled in various ways, and

(ii) A number of draws from the box.

We will imagine drawing tickets from the box in one of two ways.

(*) ***With replacement*** – after a ticket is drawn and observed, it is replaced in the box. In this case the composition of the box doesn't change from draw to draw, and the results of the different draws are *independent.*

(*) ***Without replacement*** – each ticket that is drawn from the box stays out of the box for the remaining draws. In this case, the composition of the box changes from draw to draw, and the results of the different draws are generally *not* independent.

Example 3. A box contains 50 tickets: 20 red tickets, 15 blue tickets, 10 green tickets and 5 orange tickets.

(*) If 3 tickets are drawn from the box at random ***with*** replacement, what is the probability that all three of the tickets are blue?

$\Rightarrow$ results of the draws are independent, so

$$\begin{aligned} &P(\text{1st blue and 2nd blue and 3rd blue}) \\ &\qquad = P(\text{1st blue}) \cdot P(\text{2nd blue}) \cdot P(\text{3rd blue}) \\ &\qquad = \frac{15}{50} \cdot \frac{15}{50} \cdot \frac{15}{50} = 2.7\% \end{aligned}$$

(*) If 3 tickets are drawn from the box at random ***without*** replacement, what is the probability that all three of the tickets are blue?

$\Rightarrow$ results of the draws are ***not*** independent, so

$$
\begin{aligned}
&P(\text{1st blue and 2nd blue and 3rd blue}) \\
&\quad = P((\text{1st blue and 2nd blue}) \text{ and 3rd blue}) \\
&\quad = P(\text{1st and 2nd blue}) \cdot P(\text{3rd blue}|\text{1st and 2nd blue}) \\
&\quad = \overbrace{P(\text{1st blue}) \cdot P(\text{2nd blue}|\text{1st blue})}^{P(\text{1st and 2nd blue})} \cdot P(\text{3rd blue}|\text{1st and 2nd blue}) \\
&\quad = \frac{15}{50} \cdot \frac{14}{49} \cdot \frac{13}{48} = 2.32\%
\end{aligned}
$$

Example 4. A box contains 1 red ticket, 2 blue tickets and 2 yellow tickets. If one ticket is drawn randomly from the box, what is the probability that it is red ***or*** yellow?

Three of the five tickets in the box are either red or yellow so

$$P(\text{red or yellow}) = 3/5 = 1/5 + 2/5 = P(\text{red}) + P(\text{yellow}).$$

Example 5. Two tickets are drawn at random, with replacement from the box above. What is the probability that the first ticket is red ***or*** the second ticket is yellow?

It is tempting to add the probabilities, as we did above:

$$\begin{aligned} P((\text{first red}) \text{ or } (\text{second yellow})) &= P(\text{first red}) + P(\text{second yellow}) \\ &= 1/5 + 2/5 \\ &= 3/5 = 60\%. \end{aligned}$$

But this would be wrong in this case...

The table below lists all the possible pairs of tickets that we can draw (with replacement) from our box of five tickets. In this table, $y1$ and $y2$ are the first and second yellow tickets in the box, and $b1$ and $b2$ are the first and second blue tickets.

(r,r)	(r,b1)	(r,b2)	(r,y1)	(r,y2)
(b1,r)	(b1,b1)	(b1,b2)	(b1,y1)	(b1,y2)
(b2,r)	(b2,b1)	(b2,b2)	(b2,y1)	(b2,y2)
(y1,r)	(y1,b1)	(y1,b2)	(y1,y1)	(y1,y2)
(y2,r)	(y2,b1)	(y2,b2)	(y2,y1)	(y2,y2)

($\star$) The outcomes listed above are all equally likely (why?).

($\star$) The first *row* of the table includes all pairs where the *first* ticket is red.

($\star$) The last two *columns* include all pairs where the *second* ticket is yellow.

($\star$) 13 of the 25 pairs of draws have the feature we want (first-ticket-red ***or*** second-ticket-yellow), so

$$P((\text{first-ticket-red}) \text{ or } (\text{second-ticket-yellow})) = 13/25 = 52\%.$$

What is the difference between the two examples...?

The two events in the first example are ***mutually exclusive***: if the (single) ticket we draw is red, then it cannot be yellow and vice versa.

The two events in the second example are ***not mutually exclusive***. It is possible that the first ticket will be red *and* the second ticket will be yellow. In fact the chance that this happens is $2/25 = 8\%$.

Simply adding the chances in the second example, as in the first example, overestimates the probability. The pairs *first-ticket-red* ***and*** *second-ticket-yellow* are counted twice, and the chance of these pairs should be subtracted from the sum of the probabilities to get the right answer:

$$P((\text{first-red})\ \textbf{or}\ (\text{second-yellow})) = 1/5 + 2/5 - 2/25 = 13/25 = 52\%.$$

Simple rule to remember: *If the events E and F are mutually exclusive, which means that $P(E\ \textit{and}\ F) = 0$, then*

$$P(E\ \textit{or}\ F) = P(E) + P(F).$$

Example 6. Suppose that a fair coin is tossed 4 times...

(*) In this context, *tossing a fair coin* is taken to mean that

$$P(\text{heads}) = P(\text{tails}) = 50\%$$

on each toss and also that the results of different tosses are *independent* of each other.

(i) *What are the chances that we observe* 0 *heads in* 4 *tosses?*

Observing 0 heads means that each of the four tosses resulted in tails, so

$$\begin{aligned} P(0 \text{ H in 4 tosses}) &= P(4 \text{ Ts in 4 tosses}) \\ &= P(\text{T on 1st and T on 2nd and T on 3rd and T on 4th}) \\ &= P(T) \cdot P(T) \cdot P(T) \cdot P(T) \\ &= \frac{1}{2} \cdot \frac{1}{2} \cdot \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{16} = 6.25\% \end{aligned}$$

(ii) *What is the probability of observing exactly* 1 *head in* 4 *tosses?*

To observe exactly 1 head in 4 tosses means that we observe one of the 4 sequences

$$HTTT,\ THTT,\ TTHT,\ TTTH.$$

In other words, “observing 1 head in 4 tosses” is the same as “observing $HTTT$ or $THTT$ or $TTHT$ or $TTTH$.”

These four sequences are all mutually exclusive of each other (seeing one of them excludes the possibility that you saw one of the others), so

$$\begin{aligned} P(1 \text{ H in 4 tosses}) &= P(HTTT) + P(THTT) + P(TTHT) + P(TTTH) \\ &= P(H)P(T)P(T)P(T) + P(H)P(T)P(T)P(T) \\ &\qquad + P(T)P(T)P(H)P(T) + P(T)P(T)P(T)P(H) \\ &= 6.25\% + 6.25\% + 6.25\% = 6.25\% = 25\% \end{aligned}$$

because

$$P(HTTT) = P(THTT) = P(TTHT) = P(TTTH) = \frac{1}{2} \cdot \frac{1}{2} \cdot \frac{1}{2} \cdot \frac{1}{2} = 6.25\%$$

(iii) *What is the probability of observing exactly* 3 *heads in* 4 *tosses?*

Observing 3 heads in 4 tosses means observing 1 tail in 4 tosses. Since the chances for heads and tails are the same, the probability of observing exactly 1 tail in 4 tosses is the same as the probability of observing 1 head in 4 tosses. We already figured this out...

$$P(3 \text{ H in } 4 \text{ tosses}) = P(1 \text{ T in } 4 \text{ tosses}) = P(1 \text{ H in } 4 \text{ tosses}) = 25\%$$

(iv) *What is the probability of observing exactly* 4 *heads in* 4 *tosses?*

Observing 4 heads in 4 tosses is the same as observing 0 tails in 4 tosses, and this has the same probability as observing 0 heads in 4 tosses...

$$P(4 \text{ H in } 4 \text{ tosses}) = P(0 \text{ T in } 4 \text{ tosses}) = P(0 \text{ H in } 4 \text{ tosses}) = 6.25\%$$

You may have noticed that I skipped the question about 2 heads in 4 tosses...

(v) *What is the probability of observing 2 heads in 4 tosses?*

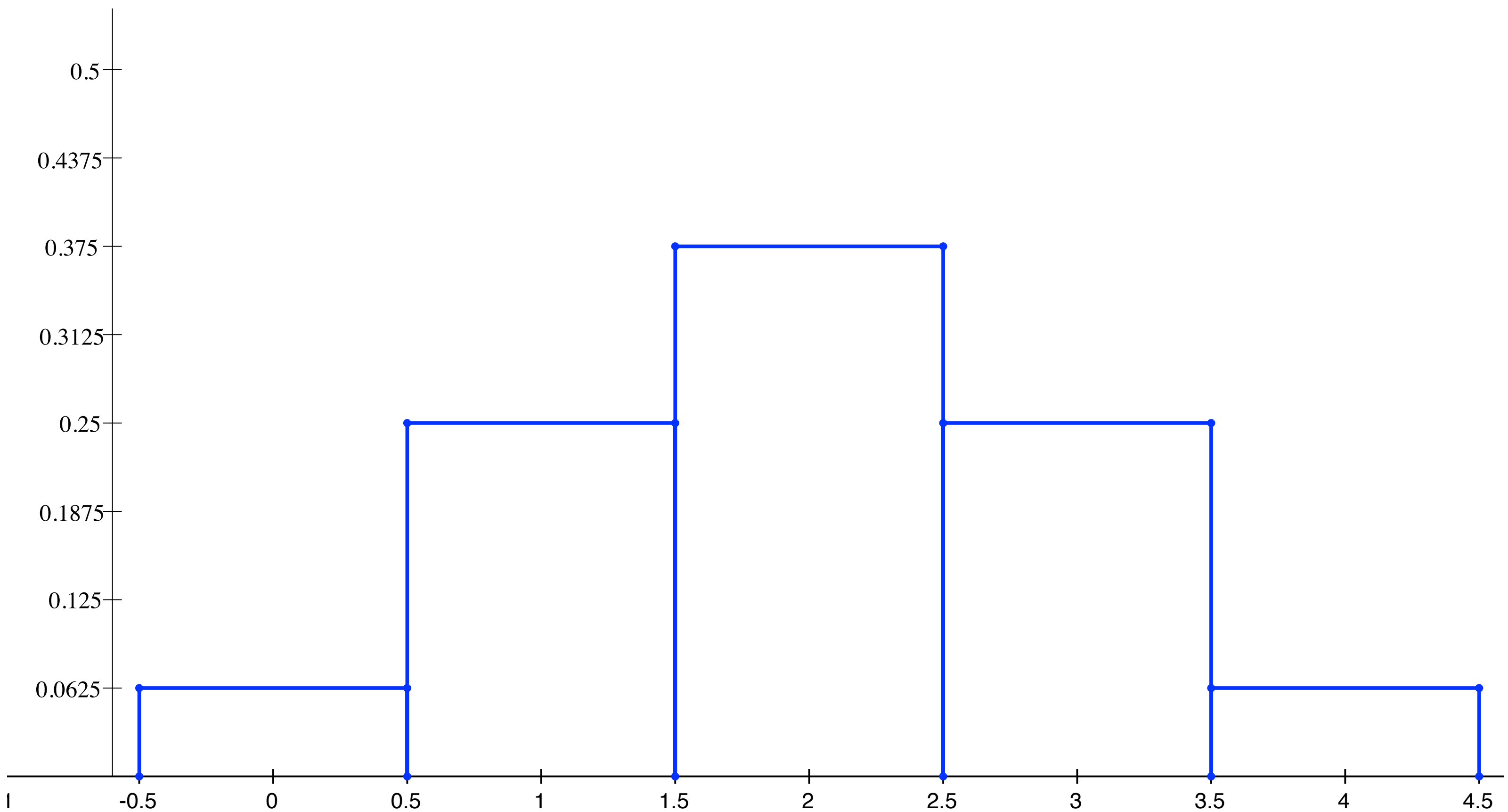
When you toss a coin 4 times, you observe either 0 heads, 1 head, 2 heads, 3 heads or 4 heads. Moreover, these events are all mutually exclusive (you can't observe both 3 heads and 0 heads, for example.

This means that the event $E =$'not 2 heads in 4 tosses' is the same as the event '0 heads, 1 head, 3 heads or 4 heads in 4 tosses'. So we can find the probability of 2 heads in 4 tosses, using the rule for *complements*...

$$
\begin{aligned}
P(\text{2 heads in 4 tosses}) &= 1 - P(\text{not 2 heads in 4 tosses}) \\
&= 1 - P(\text{0 heads or 1 head or 3 heads or 4 heads}) \\
&= 1 - (P(\text{0 heads}) + P(\text{1 head}) + P(\text{3 heads}) + P(\text{4 heads})) \\
&\text{(because the events are mutually exclusive)} \\
&= 1 - (0.0625 + 0.25 + 0.25 + 0.0625) = 1 - 0.625 = 0.375
\end{aligned}
$$

The probabilities for each of the five number of heads in 4 tosses of a fair coin are displayed on the next slide in a *probability histogram*

Probability histogram for the possible number of heads in 4 tosses of a fair coin: the area (and height) of each bar is equal to the probability of observing the number of heads over which the bar is centered.



Practice and Review

Example 7. If 4 tickets are drawn with replacement from

| 1 | 2 | 2 | 4 | 6 |,

what are the chances that we observe *exactly* two [2]s?

⇒ *'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.

⇒ *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

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

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 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, 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.

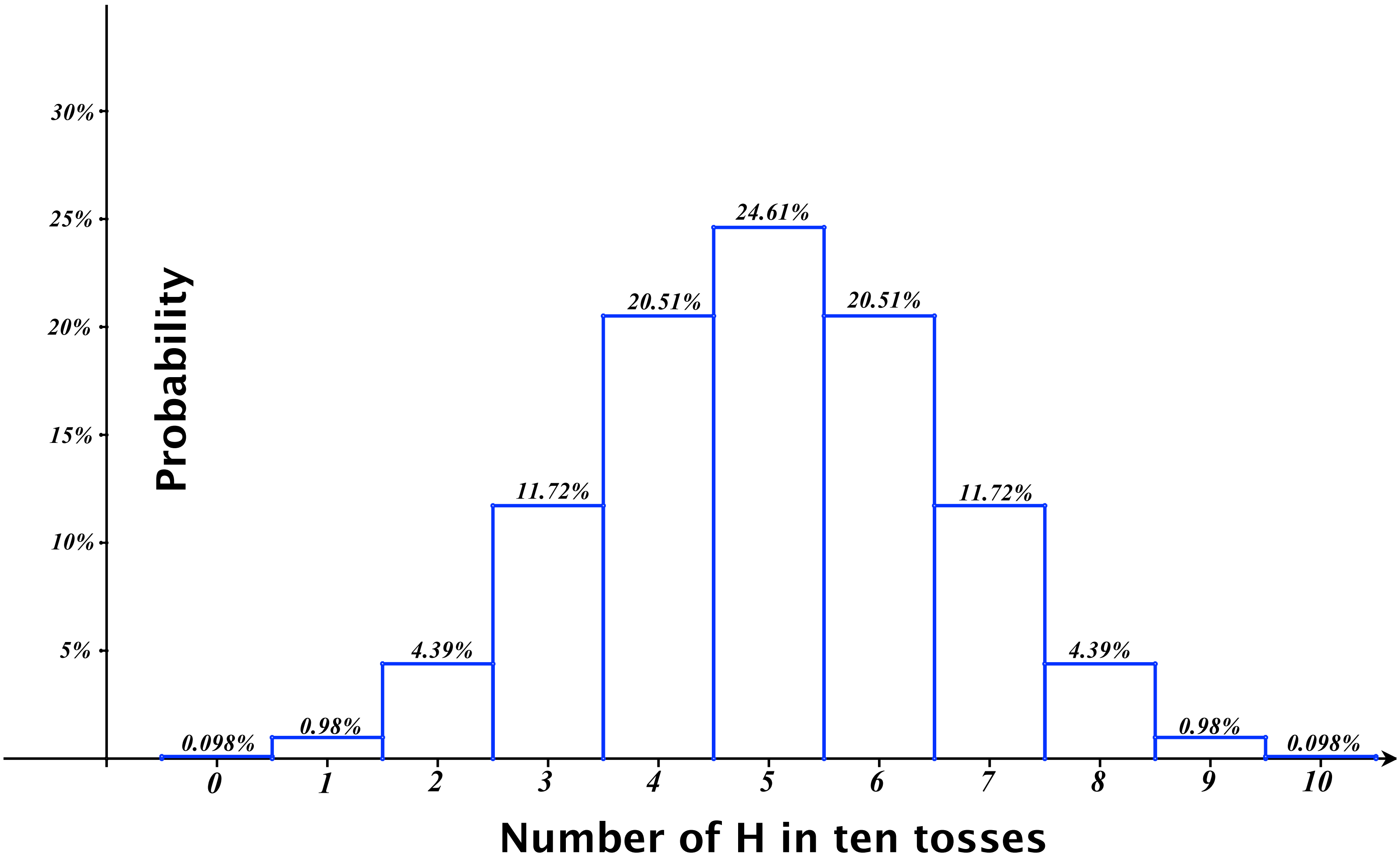
If we have a box with two tickets, for example one $\boxed{1}$ and one $\boxed{0}$, then the number of $\boxed{1}$s in n random draws with replacement from this box can be used to model the number of *heads* in n tosses of a fair coin.

(*) 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***.

$\Rightarrow$ 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...