Suppose an urn contains four balls labelled A, B, C, D. How many ways can they be selected?

\[
\begin{array}{cccc}
  \text{ABCD} & \text{BACD} & \text{CABD} & \text{DABC} \\
  \text{ACBD} & \text{BCAD} & \text{CBAD} & \text{DBAC} \\
  \text{ADBC} & \text{BDAC} & \text{CDAB} & \text{DCAB} \\
  \text{ACDB} & \text{BCDA} & \text{CBDA} & \text{DBCA} \\
  \text{ADCB} & \text{BDCA} & \text{CDBA} & \text{DCBA} \\
\end{array}
\]

There are \(4 \times 3 \times 2 \times 1 = 4! = 24\) ways to do so. In general, there are \(N!\) ways to order \(N\) items.

**Choosing \(k\) objects from \(N\)**

Suppose that two balls are drawn without replacement from an urn containing \(N\) distinctly labelled balls. There are \(N\) ways to select the first ball, and \(N - 1\) ways to select the second. Then there are \(N \times (N - 1)\) ways that two balls can be selected. Similarly, \(k\) balls can be drawn in \(N \times (N - 1) \times \cdots \times (N - k + 1)\) ways. It is more convenient to write this as \(N!/(N-k)!\):

\[
\frac{N!}{(N-k)!} = \frac{N \times (N-1) \times \cdots \times (N-k+1) \times (N-k) \times \cdots \times 2 \times 1}{(N-k) \times \cdots \times 2 \times 1} = N \times (N-1) \times \cdots \times (N-k+1).
\]

This sample can be re-ordered in \(k!\) ways. Therefore the number of distinct samples of size \(k\) drawn from \(N\) objects is

\[
\frac{N!}{(N-k)!} \times \frac{1}{k!} = \frac{N!}{k!(N-k)!} = N_C_k = \binom{N}{k}.
\]

*Based extensively on material previously taught by Eamonn Mullins.*
Example

A lottery competition involves the drawing of six numbers from a pool of 36. What is the probability of winning the lottery, i.e., correctly choosing all six?

\[ P(\text{Winning}) = \frac{1}{\binom{36}{6}} = \frac{1}{1,947,792} = 5 \times 10^{-7}. \]

Hypergeometric Distribution

Now suppose an urn contains \( N \) balls, \( K \) of which are black and \( N - k \) of which are red. If a sample of size \( n \) is selected, what is the probability that it will contain \( k \) black balls and \( n - k \) red balls? We can select \( \binom{N}{n} \) different samples of size \( n \) out of a population of \( N \). We can similarly select \( \binom{K}{k} \) different samples of red balls (of size \( k \)) and \( \binom{N-K}{n-k} \) different samples of black balls (of size \( n - k \)) from populations of size \( K \) and \( N - K \) respectively. Then the total number of different samples of \( k \) black and \( n - k \) red balls is

\[ \binom{K}{k} \binom{N-K}{n-k}, \]

and the probability that a sample of size \( n \) contains \( k \) black and \( n - k \) red balls is given by

\[ P(k, n-k) = \frac{\binom{K}{k} \binom{N-K}{n-k}}{\binom{N}{n}}. \]

This is called the hypergeometric distribution.

Example

Suppose that three balls are drawn without replacement from an urn containing three red and two black balls. The probability distribution for \( k \), the number of red balls in the sample, is then as follows:

\[
\begin{array}{ccc}
  k & P(k, 3-k) & P_k \\
  1 & \binom{3}{1} \binom{2}{2}/\binom{5}{3} & \frac{4}{10} \\
  2 & \binom{3}{2} \binom{1}{1}/\binom{5}{3} & \frac{6}{10} \\
  3 & \binom{3}{3} \binom{2}{0}/\binom{5}{3} & \frac{1}{10}
\end{array}
\]

The Birthday Problem

Suppose that \( N \) people are in a room together. What is the probability that at least two people will share a birthday? How large does \( N \) have to be before the probability is at least 50%? Let \( P(\text{No one shares a birthday}) = p_1 \). Then

\[
p_1 = \frac{365 \times 365 - 1 \times \cdots \times 365 - N + 1}{365^N}.
\]

\[
= \frac{365 \times (365 - 1) \times \cdots \times (365 - N + 1)}{365^N}.
\]

2
Then $P(\text{At least two people share a birthday}) = 1 - p_1$. Inspecting this for a range of $N$, we see that:

<table>
<thead>
<tr>
<th>$N$</th>
<th>$1 - p_1$</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>0.12</td>
</tr>
<tr>
<td>20</td>
<td>0.41</td>
</tr>
<tr>
<td>23</td>
<td>0.51</td>
</tr>
<tr>
<td>30</td>
<td>0.71</td>
</tr>
<tr>
<td>50</td>
<td>0.97</td>
</tr>
</tbody>
</table>

Now suppose that you are in a room with $N$ other people. What is the probability that at least one other person in the room shares your birthday? Let $P(\text{No one shares your birthday}) = p_2$. Then

$$p_2 = \frac{364}{365} \times \frac{364}{365} \times \cdots \times \frac{364}{365} = \left(\frac{364}{365}\right)^N,$$

and $P(\text{At least one other person shares your birthday}) = 1 - p_2$. Inspecting this probability over a range of $N$, we see that:

<table>
<thead>
<tr>
<th>$N$</th>
<th>$1 - p_2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>0.03</td>
</tr>
<tr>
<td>20</td>
<td>0.05</td>
</tr>
<tr>
<td>50</td>
<td>0.13</td>
</tr>
<tr>
<td>100</td>
<td>0.24</td>
</tr>
<tr>
<td>250</td>
<td>0.50</td>
</tr>
</tbody>
</table>

Why does this probability increase so much more slowly? In broad terms, in the first scenario, as $N$ increases we start to run out of different birthdays. In the second, the number of available days per person does not change, even as $N$ increases. Statistically, we say that this difference is caused by sampling with (for $p_2$) or without (for $p_1$) replacement.