Random Variable.
- assigns a number to each mutually exclusive event

Ex. Roll a die
RV is the result of the roll
\( X \in \{1, 2, 3, 4, 5, 6\} \)

Flip two coins
\( Y \) be # heads.
\( Y \in \{0, 1, 2\} \)

Probability Distribution.
gives the probability that
a RV takes each of the possible values.
For $X$:

<table>
<thead>
<tr>
<th>Roll</th>
<th>Probability</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>$\frac{1}{6}$</td>
</tr>
<tr>
<td>2</td>
<td>$\frac{1}{6}$</td>
</tr>
<tr>
<td>3</td>
<td>$\frac{1}{6}$</td>
</tr>
<tr>
<td>4</td>
<td>$\frac{1}{6}$</td>
</tr>
<tr>
<td>5</td>
<td>$\frac{1}{6}$</td>
</tr>
<tr>
<td>6</td>
<td>$\frac{1}{6}$</td>
</tr>
</tbody>
</table>

$P(x) = \frac{1}{6}$

For $Y$:

<table>
<thead>
<tr>
<th># heads</th>
<th>Prob</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>$\frac{1}{2}$</td>
</tr>
<tr>
<td>1</td>
<td>$\frac{1}{4}$</td>
</tr>
<tr>
<td>2</td>
<td>$\frac{1}{4}$</td>
</tr>
</tbody>
</table>

Mutually exclusive events.

Probabilities sum to 1.

Probability distributions can be defined over

discrete or continuous random variables

- e.g. Binomial
- e.g. Normal
Binomial Distribution.

Bag contains one red ball
nine green balls

Make 5 draws with replacement

What is \( P(\text{exactly 2 reds in the 5 draws}) \)?

\[
\frac{1}{10} \times \frac{1}{10} \times \frac{9}{10} \times \frac{9}{10} \times \frac{9}{10} = \left(\frac{1}{10}\right)^2 \times \left(\frac{9}{10}\right)^3
\]

\( \rightarrow \)

\( \frac{1}{10} \times \frac{1}{10} \times \frac{9}{10} \times \frac{9}{10} \times \frac{9}{10} \)

\( \left(\frac{1}{10}\right)^2 \times \left(\frac{9}{10}\right)^3 \)

\( \left(\frac{1}{10}\right)^2 \times \left(\frac{9}{10}\right)^3 \)

These possibilities are mutually exclusive.

\( \rightarrow P(\text{one of the possibilities occurs}) \) is given

\( P(\text{exactly 2 reds in 5 draws}) \) by the addition rule

\[
= 10 \times \left(\frac{1}{10}\right)^2 \times \left(\frac{9}{10}\right)^3
\]

is there a formula that gives this number?
Binomial coefficient.

\[ n - \text{trials.} \quad \binom{n}{k} = \frac{n!}{(n-k)! \cdot k!} \]

\[ k - \text{successes.} \]

\[ n! = n \text{ factorial} \]

\[ 4! = 4 \times 3 \times 2 \times 1 \]

\[ \# \text{ways of arranging } n \text{ items.} \]

\[ 0! = 1 \]

\[ n! = n \times (n-1) \times (n-2) \ldots \times 2 \times 1 \]

\[ \binom{n}{k} = \# \text{ways to pick } k \text{ out of } n \text{ if order doesn't matter.} \]

\[ p(k) = \frac{n!}{k! \cdot (n-k)!} \cdot p^k \cdot (1-p)^{n-k} \]

\[ p - \text{probability of success on any trials.} \]
Example.

A family has 4 children.

What's the probability that they have more girls than boys?

\[ P(3 \text{ out of } 4) + P(4 \text{ out of } 4) = \]

\[ \frac{4!}{3!(4-3)!} \left( \frac{1}{2} \right)^3 \left( \frac{1}{2} \right)^1 + \frac{4!}{4!0!} \left( \frac{1}{2} \right)^4 \left( \frac{1}{2} \right)^0 \]

\[ = \frac{24}{6 \times 1} \cdot \frac{1}{8} + \frac{24}{24} \cdot \frac{1}{16} \]

\[ = \frac{1}{4} + \frac{1}{16} = \frac{5}{16} \approx 31\% . \]

In class 16 families with 4 children.

6 with more girls than boys.

\[ \frac{6}{16} = 37.5\% . \]
Conditions for the Binomial Distribution to apply:

1. Fixed number of trials, $n$.
2. Trials are independent.
3. Each trial has only two possible outcomes. ("success" or "failure").
4. The probability of success, $p$, is the same for each trial

\[
\# \text{successes, } k \sim \text{Bin}(n, p).
\]

Ex.: Flip coin twice, count # heads.

# correct guesses on multiple choice test.

Random sample of 100 students, count by gender.
Multiple choice test with 5 questions
Each with 5 options, one of which is correct.
You guess randomly.

Let \( z = \# \) guess correctly.

\[
p(z) = \frac{5!}{(5-z)!z!} \left( \frac{1}{5} \right)^z \left( \frac{4}{5} \right)^{5-z}
\]

\[
p(z=5) = \frac{5!}{0!5!} \left( \frac{1}{5} \right)^5 \left( \frac{4}{5} \right)^0 = 0.0003
\]

\[
p(z=3) = \frac{5!}{2!3!} \left( \frac{1}{5} \right)^3 \left( \frac{4}{5} \right)^2 = 0.0512
\]

<table>
<thead>
<tr>
<th>( z )</th>
<th>( p(z) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0.328</td>
</tr>
<tr>
<td>1</td>
<td>0.410</td>
</tr>
<tr>
<td>2</td>
<td>0.205</td>
</tr>
<tr>
<td>3</td>
<td>0.051</td>
</tr>
<tr>
<td>4</td>
<td>0.006</td>
</tr>
<tr>
<td>5</td>
<td>0</td>
</tr>
</tbody>
</table>

\[
p(\text{at least 2 right}) = 1 - p(0 \text{ right}) - p(1 \text{ right})
\]

\[
= 1 - 0.328 - 0.410
\]

\[
= 0.262.
\]
A probability distribution has a theoretical mean + standard deviation.

The mean of a probability distribution is called the expected value.

\[ \mu = E[\omega] = \sum [\omega \cdot p(\omega)] \]

"the expectation of \( \omega \)"

\[ E[3] = 0 \times 0.328 + 1 \times 0.410 + 2 \times 0.205 + 3 \times 0.051 + 4 \times 0.006 + 5 \times 0.0 = 1 \]
\[ Y = \pm 1 \] are 2 coins.

\[ E[Y] = 0 \times \frac{1}{4} + 1 \times \frac{1}{2} + 2 \times \frac{1}{4} = 1 \]

\[ X = \text{outcome of die roll} \]

\[ E[X] = 1 \times \frac{1}{6} + 2 \times \frac{1}{6} + 3 \times \frac{1}{6} + 4 \times \frac{1}{6} + 5 \times \frac{1}{6} + 6 \times \frac{1}{6} \]

\[ = 3 \frac{1}{2} \]

\[ \sigma^2 = \text{Var}[X] = \sum [(x - \mu)^2 \cdot p(x)] = E[(x - \mu)^2] \]

\[ \sigma = \sqrt{\sum [(x - \mu)^2 \cdot p(x)]} \]

\[ \sigma^2 = E[x^2] - (E[x])^2. \quad \text{← sometimes easier to compute} \]

For a Binomial

\[ \mu = np \]

\[ \sigma = \sqrt{np(1-p)} \]

\[ Y = \# H \text{ or } 2 \text{ coins} \]

\[ \sigma_Y = \sqrt{2 \times \frac{1}{2} \times \frac{1}{2}} = \frac{1}{\sqrt{2}} \]

\[ E - \# \text{ correct guesses} \]

\[ \sigma_2 = \sqrt{5 \times \frac{1}{5} \times \frac{4}{5}} = \frac{2}{\sqrt{5}} = 0.89. \]
Recall: typically 95.0% of data lie in range \( \mu \pm 2\sigma \).

\[ Z: \mu = 1 \quad \mu + 2\sigma = 2.79. \]
\[ \sigma = 0.89 \]

If a student gets \( >3 \) correct, it starts to suggest that something other than guessing is going on.

**Poisson Distribution.**

When events occur randomly at a certain rate, the \# of events that occur in an interval of time (or space) has a Poisson distribution.
Examples.

# injuries
# typos in a paper.
# earthquakes in a year.
# chips in a cookie

\[ p(x) = \frac{\mu^x e^{-\mu}}{x!} \] 

\( \mu \) = 2.71828...

Mean: \( \mu \)
Variance: \( \mu \)
Std. dev: \( \sqrt{\mu} \)

Example: suppose you typically make 4 typos per page.
You type a 9-page paper.
Would it be unusual to make only 20 typos?

For 1 page, \( x \), \( \mu x = 4 \)
9. \( \mu y = 4 \times 9 = 36 \)
\( \sigma_y = \sqrt{36} = 6 \)
Empirical rule: consider $\mu - 2\sigma$.

$36 - 2 \times 6$

$24.$

$20 \leq 24$ so this would be considered unusual.
You arrive in one of the long intervals and chances are there is a clump after the interval (otherwise there would not have been a long interval for you to arrive in).