Note: The basic building block of probability is set theory:
Suppose we have a well defined sample space $S$ (a well defined universe $U$ in set theory
‘lingo’) and events $E_1$, $E_2$, etc. (sets $E_1$, $E_2$, etc.), yada, yada, yada. The basic ideas are
grounded in the sets!

Definition 1: If $X$ is a discrete random variable, the function given by $f(x) = \Pr(X = x)$ for each
$x$ in the domain of the function is called the probability mass function (p. m. f.).

Theorem 1: A function serves as a p. m. f. of a discrete random variable iff its values $f(x)$
satisfy both:
1. $f(x) \geq 0 \quad \forall x \in \text{dom}(f)$ and 2. $\sum_{x} f(x) = 1$.

Definition 2: If $X$ is a discrete random variable and the function given by $f(x) = \Pr(X = x)$ for
each $x$ in the domain of the function is the p. m. f. at $x$, then the expected value (or mean) of $X$
is $E[X] = \sum_{x} x \cdot f(x)$. $E[X] = \mu$

Definition 3: If $X$ is a discrete random variable and the function given by $f(x) = \Pr(X = x)$ for
each $x$ in the domain of the function is the p. m. f. at $x$, then the $r^{th}$ moment about the origin of
$X$ is $E[X^r] = \sum_{x} x^r \cdot f(x)$. $E[X^r] = \mu_r'$

Definition 4: If $X$ is a discrete random variable and the function given by $f(x) = \Pr(X = x)$ for
each $x$ in the domain of the function is the p. m. f. at $x$, then the variance (or second moment
about the mean) of $X$ is $\text{Var}[X] = \sum_{x} (x - E[X])^2 \cdot f(x)$.

$\text{Var}[X] = \sigma^2$ $\text{Var}[X] = E[(X - \mu)^2]$

Definition 5: If $X$ is a discrete random variable and the function given by $f(x) = \Pr(X = x)$ for
each $x$ in the domain of the function is the p. m. f. at $x$, then the standard deviation of $X$ is
$\text{SD}[X] = \sqrt{\sum_{x} (x - E[X])^2 \cdot f(x)}$. $\text{SD}[X] = \sigma$

Definition 6: If $X$ is a discrete random variable and the function given by $f(x) = \Pr(X = x)$ for
each $x$ in the domain of the function is the p. m. f. at $x$, then the $r^{th}$ moment about the mean of
$X$ is $E[(X - \mu)^r] = \sum_{x} (x - E[X])^r \cdot f(x)$. $E[(X - \mu)^r] = \mu_r'$
Theorem 2: If $X$ is a discrete random variable and the function given by $f(x) = \Pr(X = x)$ for each $x$ in the domain of the function is the p.m.f. at $x$, then $\text{Var}[X] = \mu_2 - \mu^2 = E[X^2] - (E[X])^2$

The Bernoulli trial is a probabilistic (or stochastic) experiment that can have one of two outcomes, success ($X = 1$) or failure ($X = 0$) in which the probability of success is $p$. The parameter is $p$.

$p \in (0, 1)$
$x \in \{0, 1\}$

$\text{Ber}(x, p) = \Pr(X = x) = \begin{cases} p, & x = 1 \\ 1 - p, & x = 0 \\ 0, & \text{else} \end{cases}$

$E[X] = \mu = p$
$
\mu' = p \quad \forall \ r \in \mathbb{N}$

$\text{Var}[X] = \sigma^2 = p(1 - p)$

$\text{SD}[X] = \sigma = \sqrt{p(1 - p)}$

The Binomial variate is the number of successes in $n$-independent Bernoulli trials where the probability of success at each trial is $p$. The parameters are $p$ and $n$ (the number of trials).

$p \in (0, 1)$

$n \in \mathbb{N}$

$x \in \{0, 1, 2, \ldots, (n - 1), n\}$

$\text{Bin}(x, p, n) = \Pr(X = x) = \binom{n}{x} p^x (1 - p)^{n-x} \quad x = 0, 1, 2, \ldots, n$

$\mu = np$

$\mu' = np(np + (1 - p))$

$\mu_3 = np((n-1)(n-2)p^2 + 3p(n - 1) + 1)$

$\sigma^2 = np(1 - p)$

$\mu_3 = np(1-p)((1 - p) - p)$

$\mu_4 = np((1 + 3p(1-p)(n - 2))$
Examples of Binomial random variables:

1. Flip a balanced coin 5 times and record the number of heads. Let \( X \) be the number of heads obtained. Then

\[
X \sim j(x) \text{ where } j(x) \text{ a probability mass function such that } j(x) = \begin{cases} 
5 \cdot \left( \frac{1}{2} \right)^x \cdot \left( \frac{1}{2} \right)^{5-x} & x \in \mathbb{N}_5^* \\
0 & \text{else}
\end{cases}
\]

<table>
<thead>
<tr>
<th>( X = x )</th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>else</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \Pr(X = x) )</td>
<td>\frac{1}{32}</td>
<td>\frac{5}{32}</td>
<td>\frac{10}{32}</td>
<td>\frac{10}{32}</td>
<td>\frac{5}{32}</td>
<td>\frac{1}{32}</td>
<td>0</td>
</tr>
</tbody>
</table>

2. Flip an unbalanced coin 5 times and record the number of heads.

So, \( X \sim f(x) \) where \( f(x) \) a probability mass function such that

\[
f(x) = \begin{cases} 
5 \cdot \left( \frac{1}{3} \right)^x \cdot \left( \frac{2}{3} \right)^{5-x} & x \in \mathbb{N}_5^* \\
0 & \text{else}
\end{cases}
\]

<table>
<thead>
<tr>
<th>( X = x )</th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>else</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \Pr(X = x) )</td>
<td>\frac{1}{243}</td>
<td>\frac{10}{243}</td>
<td>\frac{40}{243}</td>
<td>\frac{80}{243}</td>
<td>\frac{80}{243}</td>
<td>\frac{32}{243}</td>
<td>0</td>
</tr>
</tbody>
</table>

3. A student takes a multiple choice test with 12 question; each question has five options for a response; each question has one and only one option correct; and, a response on a question has no effect on a response on any or all of the other 11 questions (all questions are statistically independent, pair-wise, three-wise, etc, 12-wise)

So, \( X \sim g(x) \) where \( g(x) \) a probability mass function such that

\[
g(x) = \begin{cases} 
12 \cdot \left( \frac{1}{5} \right)^x \cdot \left( \frac{4}{5} \right)^{12-x} & x \in \mathbb{N}_{12}^* \\
0 & \text{else}
\end{cases}
\]

<table>
<thead>
<tr>
<th>( X = x )</th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
<th>12</th>
<th>else</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \Pr(X = x) )</td>
<td>\frac{16777216}{244140625}</td>
<td>\frac{5031648}{244140625}</td>
<td>\frac{69206016}{244140625}</td>
<td>\frac{57671680}{244140625}</td>
<td>\frac{32440320}{244140625}</td>
<td>\frac{12080128}{244140625}</td>
<td>\frac{3784704}{244140625}</td>
<td>\frac{811008}{244140625}</td>
<td>\frac{126720}{244140625}</td>
<td>\frac{14080}{244140625}</td>
<td>\frac{1056}{244140625}</td>
<td>\frac{48}{244140625}</td>
<td>\frac{1}{244140625}</td>
<td>0</td>
</tr>
</tbody>
</table>