Given an experiment with sample space $S$, a **random variable** (‘**RV’**) is a real number which depends on the outcome of the experiment.

If we only care about the value taken by an RV $X$, then we can use a reduced description of the experiment which deals only with events such as

$$\{a < X \leq b\} \text{ for real values } a < b$$

and their probabilities. A convenient way to express these is in terms of the CDF of $X$:

$$F(a) = P\{X \leq a\}, \quad -\infty < a < \infty.$$
A RV $X$ is **discrete** if all of the values it can possibly take may be written as a (finite or infinite) list $x_1, x_2, \ldots$.

For a discrete RV, the probabilities of all the associated events may be expressed in terms of its PMF:

$$ p(a) = P\{X = a\}, \quad -\infty < a < \infty. $$

For any other set $A$ of real numbers, we obtain

$$ P\{X \text{ takes a value in } A\} = \sum_{a \text{ in } A \text{ such that } p(a) > 0} p(a). $$
If $X$ is a discrete RV with possible values $x_1, x_2, \ldots$, and $p$ is its PMF, then its expectation is

$$E[X] = \sum_i p(x_i)x_i = \sum_{a \text{ such that } p(a)>0} p(a)a.$$ 

This may be seen as a ‘weighted average’ of the possible values of $X$, around which those values are distributed.

TODAY we will mostly develop more tools for computing expectations, and some of their consequences.
In spite of the intuitive connection to the ‘mean’ of a bunch of statistics, sometimes calculations of expectations have to be treated with care.

Example (Ross E.g. 4.3d)

A school class of 120 students travels in 3 buses. The first bus contains 36 students, the second 40, and the third 44. When the buses arrive, one student is chosen at random. Let $X$ be the number of students on the bus that they traveled on. Find $E[X]$. 
The next example concerns a special kind of random variable which is sometimes very useful. For any event \(E \subset S\), its indicator variable \(I\) is defined by

\[
I = \begin{cases} 
1 & \text{if } E \text{ occurs} \\
0 & \text{if } \overline{E} \text{ occurs.}
\end{cases}
\]

Thus, \(I\) returns a 1 or a 0 to indicate whether or not the event \(E\) occurs.

**Example (Ross E.g. 4.3b)**

*Find \(E[I]\).*
Suppose that $X$ is a RV, and also that $g$ is some function from real numbers to real numbers. Then we may define a new RV $g(X)$. If the outcome of the experiment is $s$, then this new RV returns the result of applying $g$ to the value that $X$ takes for the outcome $s$.

Often, we know something about $X$, and want to turn that into information about $g(X)$: most obviously, its expectation.
Example (Ross E.g. 4.4a)

Let $X$ be a RV that takes the possible values $-1$, 0 and 1 with probabilities 0.2, 0.5 and 0.3, respectively. Compute $E[X^2]$. 
We can turn the calculation from the above example into a general fact.

**Proposition (Ross Prop 4.4.1)**

If $X$ is discrete with possible values $x_1, x_2, \ldots$, and $p$ is its PMF, then

- $g(X)$ is discrete with possible values $g(x_1), g(x_2), \ldots$ (except that this list may contain REPEATS), and
- we have

$$E[g(X)] = \sum p(x_i)g(x_i).$$

**IDEA:** The only thing we have to worry about is that several *different* values $x_i$ might give the same value of $g(x_i)$. 
Here is a basic use of an expectation as a ‘representative’ value for a RV. As such, it is a natural choice for ‘the thing to maximize’ in this problem.

Example (Ross E.g. 4.4b)

A store orders umbrellas in September and then sells them until April. They sell each umbrella for \( b \) dollars, and they lose \( \ell \) dollars for each umbrella they don’t end up selling. In a given year, the number customers who want umbrellas is a random variable \( X \) with PMF \( p(i) \), \( i = 0, 1, 2, \ldots \). How many umbrellas should the store buy in September to maximize their expected profit?
Another simple consequence of the preceding proposition:

**Corollary (Ross Corollary 4.4.1)**

*If $a$ and $b$ are constants then*

\[ E[aX + b] = aE[X] + b. \]
Another important quantity in statistics is the following. If $x_1, x_2, \ldots, x_n$ is a list of statistics and $\overline{x}$ is their mean, then their ‘variance’ is the quantity

$$v = \frac{(x_1 - \overline{x})^2 + (x_2 - \overline{x})^2 + \cdots + (x_n - \overline{x})^2}{n},$$

and their standard deviation is its square root, $\sqrt{v}$.

The mean serves as a kind of ‘representative’ value for the whole list of statistics, and then the variance measures how ‘spread out’ the statistics are around that representative value.

Like the mean, variance has an important generalization to RVs. Once again we focus on discrete RVs for now.
To see the need for this generalization, consider the following RVs:

- $W = 0$ always;
- $Y = 1$ or $-1$, each with probability $1/2$;
- $Z = 100$ or $-100$, each with probability $1/2$.

Then $E[W] = E[Y] = E[Z] = 0$, but $Z$ is much more ‘spread out’ than $Y$, which is more ‘spread out’ than $W$. Variance gives a way to measure this fact.
Definition (Variance)

For any discrete RV $X$, if we let $\mu = E[X]$, then the variance of $X$ is

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

Its standard deviation is the square root:

$$\text{SD}(X) = \sqrt{\text{Var}(X)}.$$
An alternative formula for variance:

\[
\text{Var}(X) = E[X^2] - (E[X])^2.
\]

Often we can easily compute it using the rule for a function of a RV.
Example (Ross E.g. 4.5a)

*Calculate $\text{Var}(X)$ if $X$ represents the outcome when a fair die is rolled.*
The following property gives more insight into what variance means:

Proposition

Let $X$ be a discrete RV and let $\mu = E[X]$. If $\text{Var}(X) = 0$, then

$$P\{X = \mu\} = 1.$$
Sometimes, the information we care about from an experiment is described by several random variables, $X$, $Y$, . . . .

Working with many RVs at once is tricky, because they can lead to all sorts of ‘composite’ events such as

$$\{ X \leq 5 \text{ and } Y > 7 \text{ and } X^2 + Y^2 \leq 100 \},$$

or whatever. The tools for handling such things will be developed later in the course.

But it turns out that the expectation of a sum of RVs, such as $E[X + Y]$, behaves quite simply, and this greatly simplifies some calculations.
Proposition (Ross Corollary 9.2)

For random variables $X_1, X_2, \ldots, X_n$, we have

$$E[X_1 + X_2 + \cdots + X_n] = E[X_1] + E[X_2] + \cdots + E[X_n].$$

This fact is often called ‘linearity of expectation’.

This proposition is extremely general. It holds far beyond discrete RVs as well, once one defines expectation appropriately. We’ll see a version for some other RVs later in the course.

For now let’s just prove it in a special case, which covers all our immediate applications.
Suppose \( S = \{ s_1, s_2, \ldots, s_n \} \) is finite, so the distribution \( P \) is determined by the single-outcome probabilities

\[
p_1 = P(\{ s_1 \}), \ p_2 = P(\{ s_2 \}), \ldots, \ p_n = P(\{ s_n \}).
\]

**Proposition**

For each \( j = 1, 2, \ldots, n \), let \( X(s_j) \) denote the value that \( X \) takes when the outcome is \( s_j \). Then

\[
E[X] = p_1 \cdot X(s_1) + \cdots + p_n \cdot X(s_n).
\]

**IDEA:** Group together \( s_j \)'s which give the same value of \( X(s_j) \).

Using this, we can now prove the linearity of expectation.
Example (Ross E.g. 4.9c)

Find the expected value of the sum of \( n \) rolls of a fair die.

Example (Ross E.g. 4.9d)

Suppose that \( n \) tests are performed, and that test \( i \) is a success with probability \( p_i \). Find the expected number of successes.

OBSERVE: Did we assume that the tests are independent?
The previous example can be turned into the following very general and useful fact.

**Proposition (Ross equation (7.3.1))**

Let \( A_1, A_2, \ldots, A_n \) be a list of events, and let \( X \) be the RV which gives the number of these events that occur (that is, for a given outcome \( s \), \( X(s) \) is the number of these events which contain \( s \)). Then

\[
E[X] = P(A_1) + P(A_2) + \cdots + P(A_n).
\]

IDEA: If \( I_i \) is the indicator variable of \( A_i \), then

\[
X = I_1 + I_2 + \cdots + I_n.
\]
Example (Ross E.g. 7.2h)

Suppose $n$ people throw their hats into the centre of a room. The hats are mixed up, and each person randomly selects one. Find the expected number of people who get their own hat back.