- Probability space: The basic object in probability. We have a *sample space* S, which we can think of as the set of all possible outcomes of an experiment. The randomness come from a *probability function* $\mathbb{P}$, which tells us the probability of various events in the sample space.

- Our probability function needs to obey two rules; Probabilities are between 0 and 1 (with the whole space having probability 1), and the probabilities of disjoint events add[1].

- Suppose we run some random experiment. Without being told the full result, we're given partial information on what happened. We can update our probabilities based on this. We call $\mathbb{P}(A \mid B)$ the *conditional probability* of A conditioned on B, and define it by

$$\mathbb{P}(A \mid B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}.$$

- From this we get *Bayes' rule*: $\mathbb{P}(A|B) = \mathbb{P}(B|A)\mathbb{P}(A)/\mathbb{P}(B)$.

- We say that two events A and B are *independent* if $\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)$, or equivalently[2] that $\mathbb{P}(A|B) = \mathbb{P}(A)$, or $\mathbb{P}(B|A) = \mathbb{P}(B)$.

- Conditional probabilities give us ways of breaking things down into more easily workable pieces. If we have a bunch of events $A_1, A_2, \ldots, A_k$ which are disjoint and cover S, then for any other event B we get

$$\begin{aligned}\mathbb{P}(B) &= \mathbb{P}(B \cap A_1) + \cdots + \mathbb{P}(B \cap A_k) \\ &= \mathbb{P}(B|A_1)\mathbb{P}(A_1) + \cdots + \mathbb{P}(B|A_k)\mathbb{P}(A_k)\end{aligned}$$

  This is very useful if B is complicated, but can be made simpler if we know about A.

- *Random variables* are the next key piece of the puzzle. Formally, a random variable is a function from the probability space to the real numbers. The intuitive picture is of a variable whose value depends on the outcome of your experiment.

- For the purposes of this class, we distinguish random variables as being either discrete-valued or continuous-valued.

- Discrete random variables are governed by a PMF (probability mass function), which is the function

$$p_X(x) = \mathbb{P}(X = x).$$

[1] In general, $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B)$. If the events are disjoint, then their intersection has probability zero, so the last term goes away and they just add.

[2] In the conditional expressions we are implicitly assuming that the probabilities are not zero.

- Continuous random variables are governed by a PDF (probability density function), which is a function f_X such that

$$\mathbb{P}(a \leq X \leq b) = \int_a^b f_X(x)\ dx$$

- Any random variable has a CDP (cumulative distribution function), which is the function

$$F_X(x) = \mathbb{P}(X \leq x)$$

- There are many types of distibutions you should know: Bernoulli, binomial, geometric, Poisson, normal, exponential, uniform, etc.

- The expectation of a random variable is given by

$$\text{(discrete case)} \qquad \mathbb{E}[X] = \sum x p_X(x) = \sum x \mathbb{P}(X = x)$$
$$\text{(continuous case)} \qquad \mathbb{E}[X] = \int x f_X(x)\ dx$$

  This is the 'balancing point' of the distribution; it's your best guess for the value of a random variable given no additional information.

- Note that we can find expectations of related random variables like so:

$$\text{(discrete case)} \qquad \mathbb{E}[\phi(X)] = \sum \phi(x) p_X(x) = \sum \phi(x) \mathbb{P}(X = x)$$
$$\text{(continuous case)} \qquad \mathbb{E}[\phi(X)] = \int \phi(x) f_X(x)\ dx$$

- We can also define the variance of the distribution as follows:

$$\operatorname{Var} X = \mathbb{E}[(X - \mathbb{E}[X])^2] = \mathbb{E}[X^2] - \mathbb{E}[X]^2$$

- The law of total probability has a natural extension to expectations. For our disjoint events $A_1, A_2, \ldots,$ covering S, we get

$$\mathbb{E}[X] = \mathbb{E}[X|A_1]\mathbb{P}(A_1) + \mathbb{E}[X|A_2]\mathbb{P}(A_2) + \cdots$$