A *finite probability space* is a set S and a function $p : S \to R_{\geq 0}$ such that $p(s) > 0$ $(\forall s \in S)$ and $\sum_{s\in S} p(s) = 1$. We refer to S as the *sample space*, subsets of S as *events*, and p as the *probability distribution*. The probability of an event $A \subseteq S$ is $p(A) = \sum_{a\in A} p(a)$. (And $p(\emptyset) = 0$.)

Two events are *disjoint* if their intersection is empty. In general we have $p(A \cup B) + p(A \cap B) = p(A) + p(B)$, and thus for disjoint events $p(A \cup B) = p(A) + p(B)$. (The first statement follows from the principle of *inclusion - exclusion*: $|A\cup B| = |A|+|B|-|A\cap B|$.)

The probability of the intersection of two events is also known as the *joint probability*: $p(A, B) \equiv p(A \cap B)$. Note that it is symmetric: $p(A, B) = p(B, A)$. Suppose we know that one event has happened and wish to ask about another. For two events A and B, the *conditional probability* of A given B is $p(A|B) = p(A, B)/p(B)$.

Example: Suppose we flip a fair coin 3 times. Let B be the event that we have at least one H and A be the event of getting exactly 2 Hs. What is the probability of A given B? In this case, $(A \cap B) = A$, $p(A) = 3/8$, $p(B) = 7/8$, and therefore $p(A|B) = 3/7$.

Note that the definition of conditional probability also gives the formula: $p(A, B) = p(A|B)p(B)$. (For three events, we have $p(A \cap B \cap C) = p(A|B \cap C)p(B|C)p(C)$, with the obvious generalization to n events.)

We can also use conditional probabilities to find the probability of an event by breaking the sample space into disjoint pieces. If $S = S_1 \cup S_2 \ldots \cup S_n$ and all pairs S_i, S_j are disjoint, then for any event A, $p(A) = \sum_i p(A|S_i)p(S_i)$.

Example: Suppose we flip a fair coin twice. Let S_1 be the outcomes where the first flip is H and S_2 be the outcomes where the first flip is T. What is the probability of $A =$ getting 2 Hs? $p(A) = (1/2)(1/2) + (0)(1/2) = 1/4$.

Two events A and B are *independent* if $p(A, B) = p(A)p(B)$. This immediately gives: A and B are independent iff $p(A|B) = p(A)$. In addition, if $p(A, B) > p(A)p(B)$ then A and B are said to be *positively correlated*, and if $p(A, B) < p(A)p(B)$ then A and B are said to be *negatively correlated*.

Example: In the example of flipping 3 coins, $p(A|B) \neq p(A)$ and therefore these two events are not independent. Let C be the event that we get at least one H and at least one T. Let D be the event that we get at most one H. $p(C) = 6/8$, $p(D) = 4/8$, and $p(C, D) = 3/8$, and therefore events C and D are independent.

A simple formula follows from the above definitions and symmetry of the joint probability: $p(A|B)p(B) = p(A, B) = p(B, A) = p(B|A)p(A)$. The resulting relation

$$p(A|B)p(B) = p(B|A)p(A) \qquad \text{(Bayes)}$$

is frequently called "Bayes' theorem" or "Bayes' rule". In the case of sets A_i that are mutually disjoint, and with $\sum_{i=1}^{n} A_i = S$, then Bayes' rule takes the form

$$p(A_i|B) = \frac{p(B|A_i)p(A_i)}{p(B|A_1)p(A_1) + \ldots + p(B|A_n)p(A_n)} \ .$$

Example 1: Consider a casino with loaded and unloaded dice. For a loaded die, the probability of rolling a 6 is 50%: $p(6|L) = 1/2$, and $p(i|L) = 1/10$ $(i = 1, \ldots, 5)$. For a fair die the probabilities are $p(i|\overline{L}) = 1/6$ $(i = 1, \ldots, 6)$. Suppose there's a 1% probability of choosing a loaded die, $p(L) = 1/100$. If we select a die at random and roll three consecutive 6's with it, what is the posterior probability, $P(L|6,6,6)$, that it was loaded?

The probability of the die being loaded, given 3 consecutive 6's, is

$$p(L|6,6,6) = \frac{p(6,6,6|L)p(L)}{p(6,6,6)} = \frac{p(6|L)^3 p(L)}{p(6|L)^3 p(L) + p(6|\overline{L})^3 p(\overline{L})}$$
$$= \frac{(1/2)^3 \cdot (1/100)}{(1/2)^3 \cdot (1/100) + (1/6)^3 \cdot (99/100)} = \frac{3}{14} \approx .21$$

(so only a roughly 21% chance that it was loaded).

Example 2: Duchenne Muscular Dystrophy (DMD) can be regarded as a simple recessive sex-linked disease caused by a mutated X chromosome ($\overline{\mathrm{X}}$). An $\overline{\mathrm{X}}\mathrm{Y}$ male expresses the disease, whereas an $\overline{\mathrm{X}}\mathrm{X}$ female is a carrier but does not express the disease. Suppose neither of a woman's parents expresses the disease, but her brother does. Then the woman's mother must be a carrier, and the woman herself therefore has an *a priori* 50/50 chance of being a carrier, $p(C) = 1/2$. Suppose she gives birth to a healthy son (h.s.). What now is her probability of being a carrier?

Her probability of being a carrier, given a healthy son, is

$$p(C|\text{h.s.}) = \frac{p(\text{h.s.}|C)p(C)}{p(\text{h.s.})} = \frac{p(\text{h.s.}|C)p(C)}{p(\text{h.s.}|C)p(C) + p(\text{h.s.}|\overline{C})p(\overline{C})} = \frac{(1/2)\cdot(1/2)}{(1/2)\cdot(1/2) + 1\cdot(1/2)} = \frac{1}{3}$$

(Intuitively what is happening is that if she's not a carrier, then there are two ways she could have a healthy son, i.e., from either of her good X's, whereas if she's a carrier there's only one way. So the probability that she's a carrier is 1/3, given the knowledge that she's had exactly one healthy son.)

Example 3: Suppose there's a rare genetic disease that affects 1 out of a million people, $p(D) = 10^{-6}$. Suppose a screening test for this disease is 100% sensitive (i.e., is always correct if one has the disease), and 99.99% specific (i.e., has a .01% false positive rate). Is it worthwhile to be screened for this disease?

The above sensitivity and specificity imply that $p(+|D) = 1$ and $p(+|\overline{D}) = 10^{-4}$, so the probability of having the disease, given a positive test (+), is

$$p(D|+) = \frac{p(+|D)p(D)}{p(+)} = \frac{p(+|D)p(D)}{p(+|D)p(D) + p(+|\overline{D})p(\overline{D})} = \frac{1\cdot 10^{-6}}{1\cdot 10^{-6} + 10^{-4}(1-10^{-6})} \approx 10^{-2}$$

and there's little point to being screened (only once).