# Discrete Probability and Counting



A *finite probability space* is a set $S$ and a real function $p(s)$ on $S$ such that:

- $p(s) \geq 0$ , $\forall s \in S$, 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 $p(A) = \sum_{a \in A} p(a)$.

(Note that $p(\emptyset) = 0$.)

**Example**: Suppose we flip a fair coin. "Fair" implies that it is equally likely to come up $H$ (heads) or $T$ (tails), and therefore $p(H) = p(T) = 1/2$.

If we assign all elements of $S$ the same probability, as in the example above, then $p$ is the *uniform distribution.*

**Example**: Suppose we flip a biased coin where the probability of $H$ is twice the probability of $T$. Since $p(H) + p(T) = 1$, this implies $p(H) = 2/3$ and $p(T) = 1/3$.

**Example:** Suppose we flip a fair coin twice. What is the probability of getting one $H$ and one $T$? The possible outcomes are $\{HH, HT, TH, TT\}$. Two out of the possible 4 outcomes give one $H$ and one $T$, each outcome has probability $1/4$, and therefore the total probability is $1/2$

Suppose we flip a fair coin $n$ times. How many possible outcomes are there? There are two choices for each flip of the coin, so there are $2^n$ possible outcomes. Each coin flip is an independent event (a notion shortly to be made precise), so the probability of getting any one of these is $1/2^n$. Now suppose we want to know the probability of getting exactly $k$ $H$s. We need to know how many of the $2^n$ strings have exactly $k$ $H$s.

The number of ways of rearranging $k$ objects is given by

$$k! = k(k-1)(k-2)\cdots 2 \cdot 1 \ ,$$

and is read $k$ factorial. (We define $0! = 1$.) That is because there are $k$ choices for the first object, then $k-1$ choices for the second object, and so on, down to two choices for the last two objects, and a single choice for the last remaining.

Similarly, the number of ways (permutations) to choose $k$ objects from a set of $n$ objects is given by $n(n-1)\cdots(n-k+1) = n!/(n-k)!$, since there are $n$ choices for the first object down to $n-k+1$ choices for the $k^{\text{th}}$ object (after having chosen the first $k-1$ objects).

If the order in which the objects are chosen does not matter, then the number of ways (combinations) to choose $k$ objects from a set of $n$ is given by dividing the above by $k!$ (the number of ways of rearranging those $k$ objects). The number of ways to choose $k$ objects from $n$, independent of order, is thus given by:

$$\binom{n}{k} = \frac{n!}{(n-k)!k!} \ .$$

Note that $\binom{n}{k} = \binom{n}{n-k}$.

These numbers are also known as the *binomial coefficients*, because they appear in the expansion of binomials (expressions of the form $(x+y)^n$). Consider $(x+y)^2 = x^2+2xy+y^2$. The coefficients of this polynomial are $\{1,2,1\}$, i.e., the numbers $\binom{2}{0}$, $\binom{2}{1}$, $\binom{2}{2}$. In general, $(x+y)^n = \binom{n}{0}x^n + \binom{n}{1}x^{n-1}y + \binom{n}{2}x^{n-2}y^2 + \ldots + \binom{n}{n-1}xy^{n-1} + \binom{n}{n}y^n$. This is because each term contains a total of $n$ $x$'s and $y$'s, and the number of times the term $x^k y^{n-k}$ occurs in the expansion is given by the number of combinations of $n$ $x$'s and $y$'s with exactly $k$ $x$'s.

A natural example for the distinction between permutations and combinations is given by dealing from a standard 52 card deck. If we deal one card to each of four players, then the number of possibilities is $52 \cdot 51 \cdot 50 \cdot 49 = 52!/48! = 6497400$. If instead we deal four cards to a single player, then since it doesn't depend what order they're dealt, the number of distinct possibilities is $52 \cdot 51 \cdot 50 \cdot 49/4 \cdot 3 \cdot 2 \cdot 1 = 52!/48!4! = \binom{52}{4} = 270725$.

Suppose we flip a fair coin 4 times. The number of ways of getting exactly two heads is given by $\binom{4}{2} = 4!/2!2! = 4{\cdot}3/2 = 6$: $E = \{HHTT,\ HTHT,\ HTTH,\ THHT,\ THTH,\ TTHH\}$. The set $S$ of all possibilities has size $|S| = 2^4 = 16$, so $p(E) = \sum_{a\in E} p(a) = |E|/|S|$ =6/16=3/8.

**Example**: Suppose we flip a fair coin 10 times. What is the probability of getting exactly 4 $H$s? First we compute $\binom{10}{4} = 210$. Then we compute the total number of outcomes $2^{10} = 1024$. Therefore the probability of getting exactly 4 $H$s is $210/1024 \approx .205$.

Similarly, the probability of $k$ $H$'s in $N$ flips is $\binom{N}{k}/2^N$. In slightly more generality, if the coin is biased, with probability $p$ for H (and hence probability $1-p$ for T), then the probability of $k$ $H$s in $N$ flips is $\binom{N}{k}p^k(1-p)^{N-k}$, where $p^k$ is the probability of $k$ $H$s, $(1-p)^{N-k}$ is the probability of $N-k$ $T$s, and $\binom{N}{k}$ counts the number of ways that $k$ $H$s can be distributed among the $N$ flips.

Two events are *disjoint* if their intersection is empty.

**Example**: In the example of flipping 2 coins, the event $A =$ 'getting exactly one $H$' and the event $B =$ 'getting exactly 2$H$s' are disjoint. But, $A$ is not disjoint from the event $C =$ 'getting exactly one $T$'. In fact, events $A$ and $C$ are the same in this case.

In general we have: $p(A\cup B) + p(A\cap B) = p(A) + p(B)$. Therefore, for disjoint events we have: $p(A\cup B) = p(A) + p(B)$. The first statement follows from the principle of *inclusion - exclusion*, which states that $|A\cup B| = |A| + |B| - |A\cap B|$.

**Example**: Say we flip a coin 10 times. What is the probability that the first flip is a $T$ or the last flip is a $T$? The number of outcomes with the first flip $T$ is $2^9$. The number of outcomes where the last flip is a $T$ is $2^9$. The number of strings with both properties is $2^8$. Hence, the number of strings with either property is $2^9 + 2^9 - 2^8 = 768$, and the probability of first or last $T$ is $768/1024 = .75$.

Two more examples given in class:

**1)** If we roll four dice, what is the probability of at least one six? a) Consider the complement problem: there is a 5/6 probability of not rolling a six for any given die, and since the four dice are independent, the probability of not rolling a six is $(5/6)^4 = 5^4/6^4 = 625/1296$. The probability of rolling at least one six is therefore $1 - 625/1296 = 671/1296 \approx .517$. b) Alternatively, recall that the number of ways of choosing $r$ objects from a collection of $N$ is $\binom{N}{r} = N!/r!(N-r)!$ .

Any of the four dice can be the one that comes up six, and the other three don't, so the number of ways that exactly one of the four dice is six is $\binom{4}{1} \cdot 5^3 = 4 \cdot 5^3 = 500$

exactly two sixes: $\binom{4}{2} \cdot 5^2 = (4 \cdot 3/2) \cdot 5^2 = 150$

exactly three sixes: $\binom{4}{3} \cdot 5 = 4 \cdot 5 = 20$

exactly four sixes: $\binom{4}{4} = 1$

The total number of possibilities is $500 + 150 + 20 + 1 = 671$, and hence the probability is $671/6^4$, in agreement with the above.

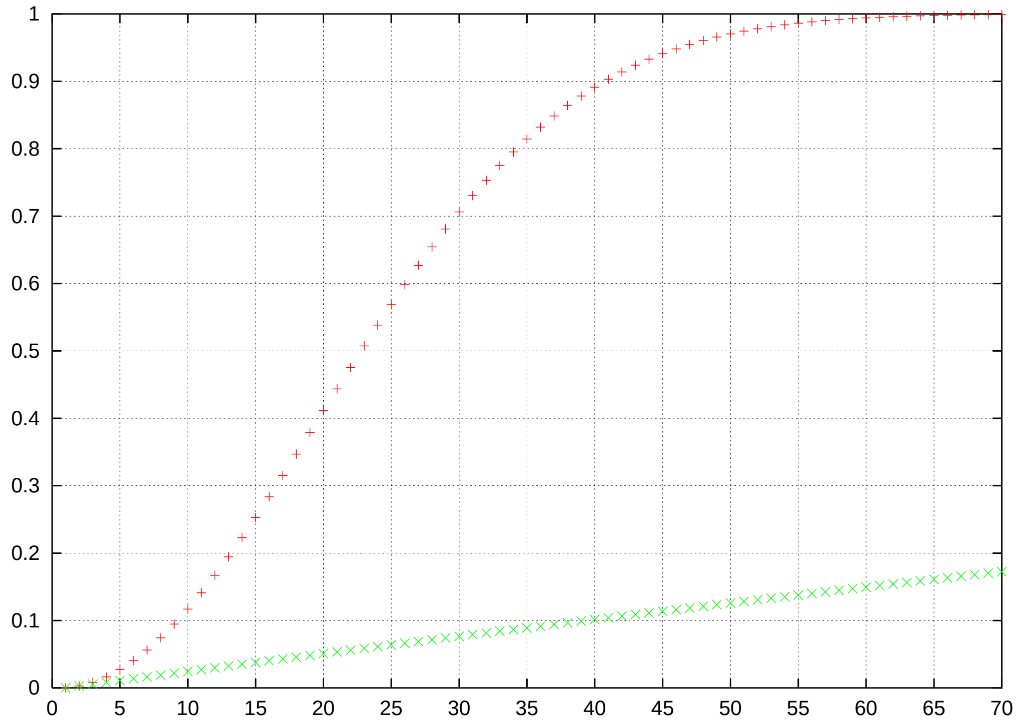
**2)** a) What is the probability that in a group of $N$ people, at least two have the same birthday?
(Simplifications: assume no leap years, and assume that all birthdays are equally likely.)

Again consider the complement problem, the probability that no two birthdays coincide. The total number of possibilities with no coincidences is $365 \cdot 364 \cdot \ldots \cdot (366-n)$ (i.e., $n$ factors each successive one with one fewer choice of day). The total number of possibilities for $n$ choices of birthdays is $365^n$, so the probability of no coincidences is $365 \cdot 364 \cdot \ldots \cdot (366-n)/365^n$. The probability that at least two coincide is therefore $1 - 365 \cdot 364 \cdot \ldots \cdot (366-n)/365^n$.

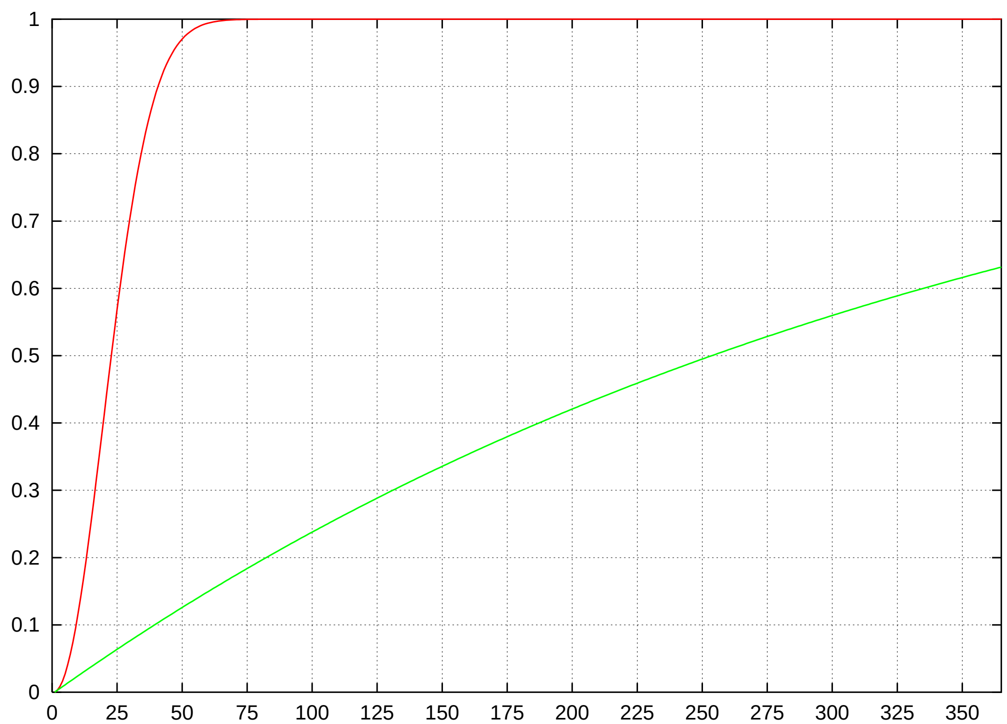
This probability is rapidly increasing as a function of $n$ and turns out to be greater than .5 for $n = 23$. (See graph on next page). For small $n$, we can estimate the probability as $\frac{n(n-2)}{2} \cdot \frac{1}{365}$, since $\frac{n(n-2)}{2} = \binom{n}{2}$ is the number of pairs, and 1/365 is the probability that any pair has a coincident birthday. (Once $n$ is too large, the pairs can no longer be considered to be roughly independent, and starts deviating by around $n \approx 15$.)

b) In a group of 23 people, what it the probability that at least one person has a birthday coincident specifically with *yours*?

In this case, we first calculate the probability that none of the 22 others (again under the above simplifications) has a birthday coincident with a given day: $(364/365)^{22}$. The probability that at least one coincides with that day is therefore $1 - (364/365)^{22} \approx .059$, so a roughly 6% chance. This probability increases more slowly as a function of the size of the group (see graphs next page, see also ipython notebook lec4.ipynb linked from course website). Note that $1 - (364/365)^n$ is well approximated by $1 - \exp(-n/365)$ for all $n$.




Red (upper): The probability $1 - \frac{365!}{(365-n)!365^n}$ that at least two birthdays coincide within a group of $n$ people, as function of $n$.
Green (lower): The probability $1 - \left(\frac{364}{365}\right)^{n-1}$ of a birthday coinciding with yours within a group of $n$ people including you.




Same as above, expanded to show $n$ up to 365. The probability for the lower case at $n = 365$ is roughly $1 - 1/e \approx .632$.

## Conditional Probability



Suppose we know that one event has happened and we wish to ask about another.

For two events $A$ and $B$, the *joint probability* of $A$ and $B$ is defined as $p(A, B) = p(A \cap B)$, i.e., the probability of the intersection of events $A$ and $B$ in the sample space, or equivalently the probability that events $A$ and $B$ both occur.

The *conditional probability* of $A$ relative to $B$ is

$$p(A|B) = p(A \cap B)/p(B)$$

and read the probability of $A$ given $B$.

**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 $H$s. What is the probability of $A$ given $B$? In this case, $(A \cap B) = A$, $p(A) = 3/8$ (why?), $p(B) = 7/8$ (why?), and therefore $p(A|B) = 3/7$.

Notice that the definition of conditional probability also gives us the formula: $p(A \cap B) = p(A|B)p(B)$. For three events we have: $p(A \cap B \cap C) = p(A|B \cap C)p(B|C)p(C)$. (What is a general rule?)

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 $A$, $p(A) = \sum_i p(A|S_i)p(S_i) = \sum_i p(A \cap S_i)$.

**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 $H$s? $p(A) = (1/2)(1/2) + (0)(1/2) = 1/4$.

Two events $A$ and $B$ are *independent* if $p(A \cap B) = p(A)p(B)$. This immediately gives: $A$ and $B$ are independent iff $p(A|B) = p(A)$.

If $p(A \cap B) > p(A)p(B)$ then $A$ and $B$ are said to be *positively correlated*.

If $p(A \cap B) < p(A)p(B)$ then $A$ and $B$ are said to be *negatively correlated*.

**Example**: In the example above of flipping 3 coins, $p(A|B) \neq p(A)$, and therefore these two events are not independent. (Since $p(A)p(B) = (3/8)(7/8) < 3/8 = p(A, B)$, they're positively correlated.)
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 $p(C \cap D) = 3/8$. Therefore events $C$ and $D$ are independent.
Whereas $p(B)p(D) = (7/8)(1/2) > 3/8 = p(B, D)$, so the events $B$ and $D$ are negatively correlated (not surprising for "at least one $H$" and "at most one $H$").

We say events $A_1, \ldots A_n$ are *mutually independent* if for all subsets $S \subseteq \{1, \ldots, n\}$, $p(\cap_{i \in S} A_i) = \prod_i p(A_i)$. (What is an example of a set of mutually independent events?)