# Introduction to Discrete Probability

Probability theory has its origins in gambling — analyzing card games, dice, roulette wheels. Today it is an essential tool in engineering and the sciences. No less so in computer science, where its use is widespread in algorithms, systems, learning theory and artificial intelligence.

Here are some typical statements that you might see concerning probability:

- The chance of getting a flush in a 5-card poker hand is about 2 in 1000.
- The chance that a particular implementation of the primality testing algorithm outputs prime when the input is composite is at most one in a trillion.
- The average time between system failures is about 3 days.
- In this load-balancing scheme, the probability that any processor has to deal with more than 12 requests is negligible.
- There is a 30% chance of a magnitude 8.0 earthquake in Northern California before 2030.

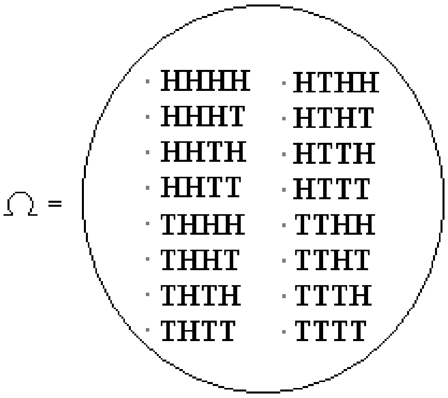
Implicit in all such statements is the notion of an underlying probability space. This may be the result of a random experiment that we have ourselves constructed (as in 1, 2 and 3 above), or some model we build of the real world (as in 4 and 5 above). None of these statements makes sense unless we specify the probability space we are talking about: for this reason, statements like 5 (which are typically made without this context) are almost content-free.

Let us try to understand all this more clearly. The first important notion here is one of a random experiment. An example of such an experiment is tossing a coin 4 times, or dealing a poker hand. In the first case an outcome of the experiment might be $HTHT$ or it might be $HHHT$. The question we are interested in might be "what is the chance that there are 2 H's." Well, there are several outcomes that would meet that condition: $HHTT, HTHT, HTTH, THHT, THTH, TTHH$. The total number of distinct outcomes to this experiment is $2^4 = 16$. If the coin is fair then all these 16 outcomes are equally likely, so the chance that there are exactly 2 H's is $3/8$.

As we saw with counting, there is a common framework in which we can view random experiments about flipping coins, dealing cards, rolling dice, etc. A finite process is the following:

We are given a finite population $U$, of cardinality $n$. In the case of coin tossing, $U = \{H, T\}$, and in card dealing, $U$ is the set of 52 cards.

An experiment consists of drawing a sample of $k$ elements from $U$. As before we will consider two cases: sampling with replacement and sampling without replacement. Thus in our coin flipping example, $n = 2$ and the sample size is $k = 4$. The outcome of the experiment is called a *sample point*. Thus $HTHT$ is an example of a sample point. The *sample space*, often denoted by $\Omega$, is the set of all possible outcomes. In our example the sample space has 16 elements.




A probability space is a sample space $\Omega$, together with a probability $\Pr[\omega]$ for each sample point $\omega$, such that

- $0 \leq \Pr[\omega] \leq 1$ for all $\omega \in \Omega$.
- $\sum_{\omega \in \Omega} \Pr[\omega] = 1$, i.e., the sum of the probabilities of all outcomes is 1.

The easiest way to assign probabilities to sample points is uniformly (as we saw earlier in the coin tossing example): if $|\Omega| = N$, then $P[x] = \frac{1}{N}\ \forall x \in \Omega$. We will see examples of non-uniform probability distributions soon.

Here's another example: dealing a poker hand. In this case, our sample space $\Omega = \{$all possible poker hands$\}$. Thus, $|\Omega| = \binom{52}{5} = \frac{52\times51\times50\times49\times48}{5\times4\times3\times2\times1} = 2,598,960$. Since the probability of each outcome is equally like, this implies that the probability of any particular hand, such as, $P[\{5h, 3c, 7s, 8c, Kh\}] = \frac{1}{2,598,960}$.

As we saw in the coin tossing example above, what we are often interested in knowing after performing an experiment is whether a certain event occurred. Thus we considered the event that there were exactly two H's in the four tosses of the coin. Here are some more examples of events we might be interested in:

- Sum of the rolls of 2 dice is $\geq 10$.
- Poker hand is a flush (i.e., all 5 cards have the same suit).
- $n$ coin tosses where $\geq \frac{n}{3}$ landed on tails.

Let us now formalize this notion of an event. Formally, an event $E$ is just a subset of the sample space, $E \subseteq \Omega$. As we saw above, the event "exactly 2 H's in four tosses of the coin" is the subset:
$\{HHTT, HTHT, HTTH, THHT, THTH, TTHH\} \subseteq \Omega$.

How should we define the probability of an event $A$? Naturally, we should just *add up* the probabilities of the sample points in $A$.

For any event $A \subseteq \Omega$, we define the probability of $A$ to be

$$\Pr[A] = \sum_{\omega \in A} \Pr[\omega].$$

Thus the probability of getting exactly two H's in four coin tosses can be calculated using this definition as follows. $A$ consists of all sequences that have exactly two H's, and so $|A| = 6$. For this example, there are

$2^4 = 16$ possible outcomes for flipping four coins. Thus, each sample point $\omega \in A$ has probability $\frac{1}{16}$, and there are 6 sample points in $A$, giving us $6 \cdot \frac{1}{16} = \frac{3}{8}$.

We will now look at examples of probability spaces and typical events that may occur in such experiments.

1. Flip a fair coin. Here $\Omega = \{H,T\}$, and $\Pr[H] = \Pr[T] = \frac{1}{2}$.

2. Flip a fair coin three times. Here $\Omega = \{(t_1,t_2,t_3) : t_i \in \{H,T\}\}$, where $t_i$ gives the outcome of the $i$th toss. Thus $\Omega$ consists of $2^3 = 8$ points, each with equal probability $\frac{1}{8}$. More generally, if we flip the coin $n$ times, we get a sample space of size $2^n$ (corresponding to all words of length $n$ over the alphabet $\{H,T\}$), each point having probability $\frac{1}{2^n}$. We can look at the event $A$ that all three coin tosses are the same. Then $A = \{HHH,TTT\}$, with each sample point having probability $\frac{1}{8}$. Thus, $\Pr[A] = \Pr[HHH] + \Pr[TTT] = \frac{1}{8} + \frac{1}{8} = \frac{1}{4}$.

3. Flip a biased coin three times. Suppose the bias is two-to-one in favor of Heads, i.e., it comes up Heads with probability $\frac{2}{3}$ and Tails with probability $\frac{1}{3}$. The sample space here is exactly the same as in the previous example. However, the probabilities are different. For example, $\Pr[HHH] = \frac{2}{3} \times \frac{2}{3} \times \frac{2}{3} = \frac{8}{27}$, while $\Pr[THH] = \frac{1}{3} \times \frac{2}{3} \times \frac{2}{3} = \frac{4}{27}$. [*Note: We have cheerfully* multiplied *probabilities here; we'll explain why this is OK later. It is* not *always OK!*] More generally, if we flip a biased coin with Heads probability $p$ (and Tails probability $1-p$) $n$ times, the probability of a given sequence is $p^r(1-p)^{n-r}$, where $r$ is the number of $H$'s in the sequence. Let $A$ be the same event as in the previous example. Then $\Pr[A] = \Pr[HHH] + \Pr[TTT] = \frac{8}{27} + \frac{1}{27} = \frac{9}{27} = \frac{1}{3}$. As a second example, let $B$ be the event that there are exactly two Heads. We know that the probability of any outcome with two Heads (and therefore one Tail) is $(\frac{2}{3})^2 \times (\frac{1}{3}) = \frac{4}{27}$. How many such outcomes are there? Well, there are $\binom{3}{2} = 3$ ways of choosing the positions of the Heads, and these choices completely specify the sequence. So $\Pr[B] = 3 \times \frac{4}{27} = \frac{4}{9}$. More generally, the probability of getting exactly $r$ Heads from $n$ tosses of a biased coin with Heads probability $p$ is $\binom{n}{r}p^r(1-p)^{n-r}$. Biased coin-tossing sequences show up in many contexts: for example, they might model the behavior of $n$ trials of a faulty system, which fails each time with probability $p$.

4. Roll two dice. Then $\Omega = \{(i,j) : 1 \le i,j \le 6\}$. Each of the 36 outcomes has equal probability, $\frac{1}{36}$. We can look at the event $A$ that the sum of the dice is at least 10, and $B$ the event that there is at least one 6. Then $\Pr[A] = \frac{6}{36} = \frac{1}{6}$, and $\Pr[B] = \frac{11}{36}$. In this example (and in 1 and 2 above), our probability space is <u>uniform</u>, i.e., all the sample points have the *same* probability (which must be $\frac{1}{|\Omega|}$, where $|\Omega|$ denotes the size of $\Omega$). In such circumstances, the probability of any event $A$ is clearly just

$$\Pr[A] = \frac{\#\text{ of sample points in } A}{\#\text{ of sample points in } \Omega} = \frac{|A|}{|\Omega|}.$$

   So for uniform spaces, computing probabilities reduces to *counting* sample points!

5. **Card Shuffling.** Shuffle a deck of cards. Here $\Omega$ consists of the 52! permutations of the deck, each with equal probability $\frac{1}{52!}$. [Note that we're really talking about an idealized mathematical model of shuffling here; in real life, there will always be a bit of bias in our shuffling. However, the mathematical model is close enough to be useful.]

6. **Poker Hands.** Shuffle a deck of cards, and then deal a poker hand. Here $\Omega$ consists of all possible five-card hands, each with equal probability (because the deck is assumed to be randomly shuffled). The number of such hands is $\binom{52}{5}$, i.e., the number of ways of choosing five cards from the deck of 52 (without worrying about the order). As we saw many lectures ago, $\binom{52}{5} = \frac{52!}{5!47!} = \frac{52\cdot51\cdot50\cdot49\cdot48}{5\cdot4\cdot3\cdot2\cdot1} = 2,598,960$. What is the probability that our poker hand is a flush (if you think about it, this is an event

since it is a subset of all possible poker hands)? [For those who are not addicted to gambling, a *flush* is a hand in which all cards have the same suit, say Hearts.] To compute this probability, we just need to figure out how many poker hands are flushes. Well, there are 13 cards in each suit, so the number of flushes in each suit is $\binom{13}{5}$. The total number of flushes is therefore $4 \cdot \binom{13}{5}$. So we have

$$\Pr[\text{hand is a flush}] = \frac{4 \cdot \binom{13}{5}}{\binom{52}{5}} = \frac{4 \cdot 13! \cdot 5! \cdot 47!}{5! \cdot 8! \cdot 52!} = \frac{4 \cdot 13 \cdot 12 \cdot 11 \cdot 10 \cdot 9}{52 \cdot 51 \cdot 50 \cdot 49 \cdot 48} \approx 0.002.$$

7. **Balls and Bins.** Throw 20 balls into 10 bins, so that each ball is equally likely to land in any bin, regardless of what happens to the other balls. Here $\Omega = \{(b_1, b_2, \ldots, b_{20}) : 1 \le b_i \le 10\}$; the component $b_i$ denotes the bin in which ball $i$ lands. There are $10^{20}$ possible outcomes (why?), each with probability $\frac{1}{10^{20}}$. More generally, if we throw $m$ balls into $n$ bins, we have a sample space of size $n^m$. [Note that example 2 above is a special case of balls and bins, with $m = 3$ and $n = 2$.] Let $A$ be the event that bin 1 is empty. Again, we just need to count how many outcomes have this property. And this is exactly the number of ways all 20 balls can fall into the remaining nine boxes, which is $9^{20}$. Hence $\Pr[A] = \frac{9^{20}}{10^{20}} = (\frac{9}{10})^{20} \approx 0.12$. What is the probability that bin 1 contains at least one ball? This is easy: this event, call it $\bar{A}$, is the *complement* of $A$, i.e., it consists of precisely those sample points that are not in $A$. So $\Pr[\bar{A}] = 1 - \Pr[A] \approx 0.88$. More generally, if we throw $m$ balls into $n$ bins, we have
$$\Pr[\text{bin 1 is empty}] = \left(\frac{n-1}{n}\right)^m = \left(1 - \frac{1}{n}\right)^m.$$
As we shall see, balls and bins is another probability space that shows up very often in Computer Science: for example, we can think of it as modeling a load balancing scheme, in which each job is sent to a random processor.

# Birthday Paradox

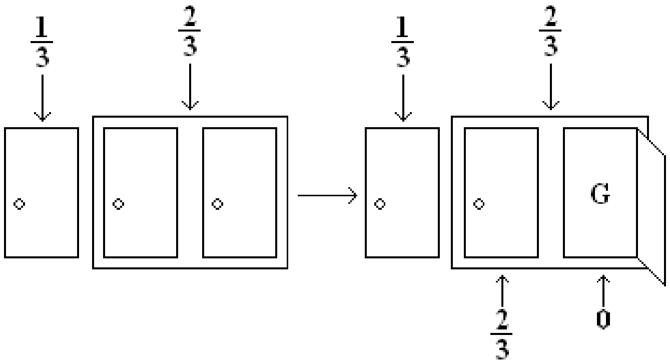
The birthday paradox is a remarkable phenomenon that examines the chances that two people in a group have the same birthday. It is a paradox not because of a logical contradiction, but because it goes against intuition. For ease of calculation, we take the number of days in a year to be 365. If we consider the case where there are $n$ people in a room, then $|\Omega| = 365^n$. Let $A$ = "At least two people have the same birthday," and let $B$ = "No two people have the same birthday." It is clear that $P[A] = 1 - P[B]$. We will calculate $P[B]$, since it is easier, and then find out $P[A]$. How many ways are there for no two people to have the same birthday? Well, there are 365 choices for the first person, 364 for the second, ..., $365 - n + 1$ choices for the $n^{th}$ person. Thus, $P[B] = \frac{|B|}{|\Omega|} = \frac{365 \times 364 \times \cdots \times (365-n+1)}{365^n}$. Then $P[A] = 1 - \frac{365 \times 364 \times \cdots \times (365-n+1)}{365^n}$. In fact, at $n = 23$ people, you should be willing to bet that at least two people do have the same birthday, since the chances of you winning are over 50%! The chances increase dramatically, and at $n = 60$ people the chances are over 99%.

# The Monty Hall Problem

In an (in)famous 1970s game show hosted by one Monty Hall, a contestant was shown three doors; behind one of the doors was a prize, and behind the other two were goats. The contestant picks a door (but doesn't open it). Then Hall's assistant (Carol), opens one of the other two doors, revealing a goat (since Carol knows where the prize is, she can always do this). The contestant is then given the option of sticking with his current door, or switching to the other unopened one. He wins the prize if and only if his chosen door is the correct one. The question, of course, is: Does the contestant have a better chance of winning if he switches doors?

Intuitively, it seems obvious that since there are only two remaining doors after the host opens one, they must have equal probability. So you may be tempted to jump to the conclusion that it should not matter whether or not the contestant stays or switches. We will see that actually, the contestant has a better chance of picking the car if he or she uses the switching strategy. We will first give an intuitive pictorial argument, and then take a more rigorous probability approach to the problem.

To see why it is in the contestant's best interests to switch, consider the following. Initially when the contestant chooses the door, he or she has a $\frac{1}{3}$ chance of picking the car. This must mean that the other doors combined have a $\frac{2}{3}$ chance of winning. But after Carol opens a door with a goat behind it, how do the probabilities change? Well, the door the contestant originally chose still has a $\frac{1}{3}$ chance of winning, and the door that Carol opened has no chance of winning. What about the last door? It must have a $\frac{2}{3}$ chance of containing the car, and so the contestant has a higher chance of winning if he or she switches doors. This argument can be summed up nicely in the following picture:




What is the sample space here? Well, we can describe the outcome of the game (up to the point where the contestant makes his final decision) using a triple of the form $(i, j, k)$, where $i, j, k \in \{1,2,3\}$. The values $i, j, k$ respectively specify the location of the prize, the initial door chosen by the contestant, and the door opened by Carol. Note that some triples are not possible: e.g., $(1,2,1)$ is not, because Carol never opens the prize door. Thinking of the sample space as a tree structure, in which first $i$ is chosen, then $j$, and finally $k$ (depending on $i$ and $j$), we see that there are exactly 12 sample points.

Assigning probabilities to the sample points here requires pinning down some assumptions:

- The prize is equally likely to be behind any of the three doors.
- Initially, the contestant is equally likely to pick any of the three doors.
- If the contestant happens to pick the prize door (so there are two possible doors for Carol to open), Carol is equally likely to pick either one.

From this, we can assign a probability to every sample point. For example, the point $(1,2,3)$ corresponds to the prize being placed behind door 1 (with probability $\frac{1}{3}$), the contestant picking door 2 (with probability $\frac{1}{3}$), and Carol opening door 3 (with probability 1, because she has no choice). So

$$\Pr[(1,2,3)] = \frac{1}{3} \times \frac{1}{3} \times 1 = \tfrac{1}{9}.$$

[Note: Again we are multiplying probabilities here, without proper justification!] Note that there are six outcomes of this type, characterized by having $i \neq j$ (and hence $k$ must be different from both). On the other hand, we have

$$\Pr[(1,1,2)] = \frac{1}{3} \times \frac{1}{3} \times \frac{1}{2} = \tfrac{1}{18}.$$

And there are six outcomes of this type, having $i = j$. These are the only possible outcomes, so we have completely defined our probability space. Just to check our arithmetic, we note that the sum of the probabilities of all outcomes is $(6 \times \frac{1}{9}) + (6 \times \frac{1}{18}) = 1$.

Let's return to the Monty Hall problem. Recall that we want to investigate the relative merits of the "sticking" strategy and the "switching" strategy. Let's suppose the contestant decides to switch doors. The event $A$ we are interested in is the event that the contestant wins. Which sample points $(i, j, k)$ are in $A$? Well, since the contestant is switching doors, his initial choice $j$ cannot be equal to the prize door, which is $i$. And all outcomes of this type correspond to a win for the contestant, because Carol must open the second non-prize door, leaving the contestant to switch to the prize door. So $A$ consists of all outcomes of the first type in our earlier analysis; recall that there are six of these, each with probability $\frac{1}{9}$. So $\Pr[A] = \frac{6}{9} = \frac{2}{3}$. That is, using the switching strategy, the contestant wins with probability $\frac{2}{3}$! It should be intuitively clear (and easy to check formally — try it!) that under the sticking strategy his probability of winning is $\frac{1}{3}$. (In this case, he is really just picking a single random door.) So by switching, the contestant actually improves his odds by a huge amount!

This is one of many examples that illustrate the importance of doing probability calculations systematically, rather than "intuitively." Recall the key steps in all our calculations:

- What is the <u>sample space</u> (i.e., the experiment and its set of possible outcomes)?
- What is the <u>probability</u> of each outcome (sample point)?
- What is the <u>event</u> we are interested in (i.e., which subset of the sample space)?
- Finally, compute the probability of the event by adding up the probabilities of the sample points inside it.

Whenever you meet a probability problem, you should always go back to these basics to avoid potential pitfalls. Even experienced researchers make mistakes when they forget to do this — witness many erroneous "proofs", submitted by mathematicians to newspapers at the time, of the fact that the switching strategy in the Monty Hall problem does not improve the odds.