EECS 70 Discrete Mathematics and Probability Theory
Spring 2020 Note 13

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 EECS, where its use is widespread in algorithms, systems, signal processing, learning theory and control/AI.

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

1. The chance of getting a flush in a 5-card poker hand is about 2 in 1000.
2. The chance that this randomized primality testing algorithm outputs "prime" when the input is not prime is at most one in a trillion.
3. In this load-balancing scheme, the probability that any processor has to deal with more than 12 requests is negligible.
4. The average time between system failures is about 3 days.
5. There is a 30% chance of a magnitude 8.0 earthquake in Northern California before 2030.

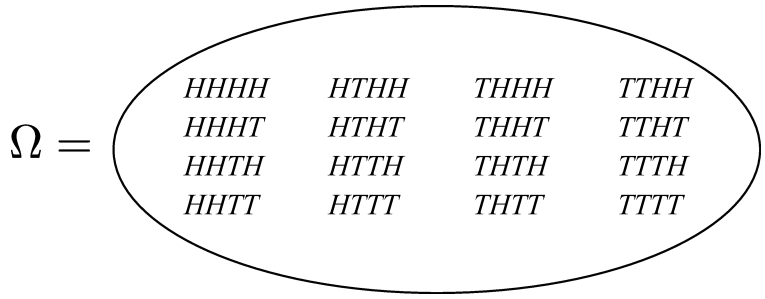
Implicit in all such statements is the notion of an underlying *probability space*. This may be the result of a random experiment that we have ourselves constructed (as in 1, 2 and 3 above), or some model we build of the real world (as in 4 and 5 above). None of these statements makes sense unless we 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.

In this note, we will try to understand all this more clearly. The first important notion here is that of a *random experiment*. We will start by introducing the space of all possible outcomes of the experiment, called a sample space. Each element of the sample space is assigned a probability which tells us how likely the outcome is to occur when we actually perform the experiment.

1 Random Experiments

In general, a random experiment consists of drawing a sample of k elements from a set S of cardinality n. The possible outcomes of such an experiment are exactly the objects that we counted in the last note. Recall from the last note that we considered four possible scenarios for counting, depending upon whether we sampled with or without replacement, and whether the order in which the k elements are chosen does or does not matter. The same will be the case for our random experiments. The outcome of a random experiment is called a *sample point*, and the *sample space* (often denoted by Ω) is the set of all possible outcomes of the experiment.

An example of such an experiment is tossing a coin 4 times. In this case, $S = \{H, T\}$ and we are drawing 4 elements with replacement. $HTHT$ is an example of a sample point and the sample space Ω has 16 elements, as illustrated in the following picture:



How do we determine the chance of each particular outcome, such as $HHTH$, of our experiment? In order to do this, we need to define the probability for each sample point, as described below.

2 Probability Spaces

A probability space is a sample space Ω, together with a probability $\mathbb{P}[\omega]$ (often also denoted as $\Pr[\omega]$) for each sample point ω, such that

- (Non-negativity): $0 \leq \mathbb{P}[\omega] \leq 1$ for all $\omega \in \Omega$.
- (Total one): $\sum_{\omega \in \Omega} \mathbb{P}[\omega] = 1$, i.e., the sum of the probabilities over all outcomes is 1.

The easiest way to assign probabilities to sample points is to do it *uniformly*: if $|\Omega| = N$, then $\mathbb{P}[\omega] = \frac{1}{N}$, $\forall \omega \in \Omega$. For example, if we toss a fair coin 4 times, each of the 16 sample points (as pictured above) is assigned probability $\frac{1}{16}$. We will see examples of non-uniform probability distributions soon.

After performing an experiment, we are often interested in knowing whether a certain event occurred. For example, we might be interested in the event that there were "exactly 2 H's in four tosses of the coin". How do we formally define the concept of an event in terms of the sample space Ω? The answer is to identify the event "exactly 2 H's in four tosses of the coin" with the set of all those outcomes in which there are exactly two H's: $\{HHTT, HTHT, HTTH, THHT, THTH, TTHH\} \subset \Omega$. Hence, formally an event A is just a subset of the sample space Ω, i.e., $A \subseteq \Omega$.

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

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

Note that $0 \leq \mathbb{P}[A] \leq 1$ for all $A \subseteq \Omega$, and $\mathbb{P}[\Omega] = 1$. The probability of getting exactly two H's in four coin tosses can be calculated using this definition as follows. Event A consists of all sequences that have exactly two H's, and so $|A| = \binom{4}{2} = 6$. 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 since there are six sample points in A, we obtain $\mathbb{P}[A] = 6 \cdot \frac{1}{16} = \frac{3}{8}$.

3 Examples

We will now look at examples of random experiments and their corresponding sample spaces, along with possible probability spaces and events.

3.1 Coin Tosses

Suppose we have a coin with $\mathbb{P}(H) = p$ and $\mathbb{P}(T) = 1 - p$, and our experiment consists of flipping the coin 4 times. The sample space Ω consists of the sixteen possible sequences of H's and T's shown in the figure on the previous page.

The probability space depends on p. If $p = \frac{1}{2}$ the probabilities are assigned uniformly; the probability of each sample point is $\frac{1}{16}$. What if $p = \frac{2}{3}$? Then the probabilities are different. For example, $\mathbb{P}[HHHH] = \frac{2}{3} \times \frac{2}{3} \times \frac{2}{3} \times \frac{2}{3} = \frac{16}{81}$, while $\mathbb{P}[TTHH] = \frac{1}{3} \times \frac{1}{3} \times \frac{2}{3} \times \frac{2}{3} = \frac{4}{81}$. (Note: We have simply multiplied probabilities here; we will explain later why this is the correct thing to do for this example. It is NOT always OK to multiply probabilities like this.)

What type of events can we consider in this setting? Let A be the event that all four coin tosses are the same. Then $A = \{HHHH, TTTT\}$. $HHHH$ has probability $\left(\frac{2}{3}\right)^4$ and $TTTT$ has probability $\left(\frac{1}{3}\right)^4$. Thus, $\mathbb{P}[A] = \mathbb{P}[HHHH] + \mathbb{P}[TTTT] = \left(\frac{2}{3}\right)^4 + \left(\frac{1}{3}\right)^4 = \frac{17}{81}$.

Next, consider the event B that there are exactly two heads. The probability of any particular outcome with two heads (such as $HTHT$) is $\left(\frac{2}{3}\right)^2\left(\frac{1}{3}\right)^2$. How many such outcomes are there? There are $\binom{4}{2} = 6$ ways of choosing the positions of the heads, and these choices completely specify the sequence. So, $\mathbb{P}[B] = 6\left(\frac{2}{3}\right)^2\left(\frac{1}{3}\right)^2 = \frac{24}{81} = \frac{8}{27}$.

More generally, if we flip the coin n times, we get a sample space Ω of cardinality 2^n. The sample points are all possible length-n sequences of H's and T's. If the coin has $\mathbb{P}(H) = p$, and if we consider any sequence of n coin flips with exactly r H's, then the probability of this sequence is $p^r(1-p)^{n-r}$.

Now consider the event C that we get exactly r H's when we flip the coin n times. This event consists of exactly $\binom{n}{r}$ sample points and each has probability $p^r(1-p)^{n-r}$. So, $\mathbb{P}[C] = \binom{n}{r}p^r(1-p)^{n-r}$.

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.

3.2 Rolling Dice

Consider rolling two fair dice. In this experiment, $\Omega = \{(i, j) : 1 \leq i, j \leq 6\}$. The probability space is uniform, i.e., all sample points have the *same* probability $\frac{1}{|\Omega|} = \frac{1}{36}$. Hence, the probability of any event A is

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

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

Now consider two events: the event A that the sum of the dice is at least 10 and the event B that there is at least one 6. By enumerating the sample points contained in each event, it can be easily shown that $|A| = 6$ and $|B| = 11$. Then, by the observation above, it follows that $\mathbb{P}[A] = \frac{6}{36} = \frac{1}{6}$ and $\mathbb{P}[B] = \frac{11}{36}$.

3.3 Card Shuffling

Consider a random experiment of shuffling a deck of standard playing cards. Here, Ω is equal to the set of the 52! permutations of the deck. We assume that the probability space is uniform. (Note that we are 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.)

3.4 Poker Hands

Here's another experiment: shuffling a deck of cards and dealing a poker hand. In this case, S is the set of 52 cards and our sample space $\Omega = \{\text{all possible poker hands}\}$, which corresponds to choosing $k = 5$ objects without replacement from a set of size $n = 52$ where order does not matter. Hence, as we saw in Note 11, $|\Omega| = \binom{52}{5} = \frac{52\times51\times50\times49\times48}{5\times4\times3\times2\times1} = 2,598,960$. Assuming that the deck is well shuffled, the probability of each outcome is equally likely and we are therefore dealing with a uniform probability space.

Let A be the event that the poker hand is a flush. (For those who are not familiar with poker, a *flush* is a hand in which all cards have the same suit, say Hearts.) Since the probability space is uniform, computing $\mathbb{P}[A]$ reduces to simply computing $|A|$, the number of poker hands that are flushes. 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 $4 \times \binom{13}{5}$. Then we have

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

3.5 Balls and Bins

In this experiment, we will throw 20 (labeled) balls into 10 (labeled) bins. Assume that each ball is equally likely to land in any bin, regardless of what happens to the other balls.

If you wish to understand this situation in terms of sampling a sequence of k elements from a set S of cardinality n: the set S consists of the 10 bins, and we are sampling with replacement $k = 20$ times. The order of sampling matters, since the balls are labeled.

The sample space Ω is equal to $\{(b_1, b_2, \ldots, b_{20}) : 1 \leq b_i \leq 10 \text{ for each } i = 1, \ldots, 20\}$, where the component b_i denotes the bin in which ball i lands. The cardinality $|\Omega|$ of the sample space is equal to 10^{20}, since each element b_i in the sequence has 10 possible choices and there are 20 elements in the sequence. More generally, if we throw m balls into n bins, we have a sample space of size n^m. The probability space is uniform; as we said earlier, each ball is equally likely to land in any bin.

Let A be the event that bin 1 is empty. Since the probability space is uniform, we simply need to count how many outcomes have this property. This is exactly the number of ways all 20 balls can fall into the remaining nine bins, which is 9^{20}. Hence, $\mathbb{P}[A] = \frac{9^{20}}{10^{20}} = \left(\frac{9}{10}\right)^{20} \approx 0.12$.

Let B be the event that bin 1 contains at least one ball. This event is the *complement* $\overline{A}$ of A, i.e., it consists of precisely those sample points which are not in A. So $\mathbb{P}[B] = 1 - \mathbb{P}[A] \approx 0.88$. More generally, if we throw m balls into n bins, we have:

$$\mathbb{P}[\text{bin 1 is empty}] = \left(\frac{n-1}{n}\right)^m = \left(1 - \frac{1}{n}\right)^m.$$

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.

It is also a more general model for problems we have previously considered. For example, flipping a fair coin 3 times is a special case in which the number of balls (m) is 3 and the number of bins (n) is 2. Rolling two dice is a special case in which $m = 2$ and $n = 6$.

3.6 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. Then $S = \{1, \ldots, 365\}$, and the random experiment consists of drawing a sample of n elements from S, where the elements are the birth dates of n people in a group. Then $|\Omega| = 365^n$. This is because each sample point is a sequence of possible birthdays for n people; so there are n points in the sequence and each point has 365 possible values.

Let A be the event that at least a pair of people have the same birthday. If we want to determine $\mathbb{P}[A]$, it might be simpler to instead compute the probability of the complement of A; i.e., $\mathbb{P}[\overline{A}]$, where $\overline{A}$ is the event that no two people have the same birthday. Since $\mathbb{P}[A] = 1 - \mathbb{P}[\overline{A}]$, we can then easily compute $\mathbb{P}[A]$.

We are again working in a uniform probability space, so we just need to determine $|\overline{A}|$. Equivalently, we are computing the number of ways for no two people to have the same birthday. There are 365 choices for the first person, 364 for the second, $\ldots$, $365 - n + 1$ choices for the n-th person, for a total of $365 \times 364 \times \cdots \times (365 - n + 1)$. This is simply an application of the First Rule of Counting from Note 11; we are sampling without replacement and the order matters.

In summary, we have

$$\mathbb{P}[\overline{A}] = \frac{|\overline{A}|}{|\Omega|} = \frac{365 \times 364 \times \cdots \times (365 - n + 1)}{365^n},$$

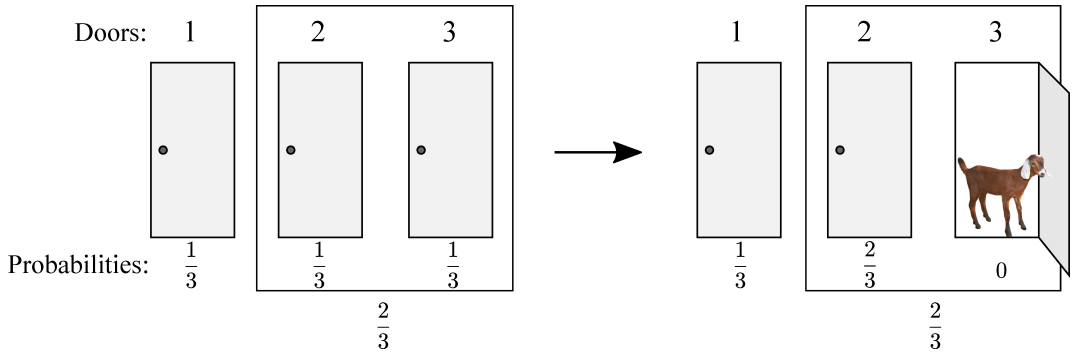
so $\mathbb{P}[A] = 1 - \mathbb{P}[\overline{A}] = 1 - \frac{365 \times 364 \times \cdots \times (365-n+1)}{365^n}$. This allows us to compute $\mathbb{P}[A]$ as a function of the number n of people. Of course, as n increases $\mathbb{P}[A]$ increases. In fact, with $n = 23$ people, you should be willing to bet that at least a pair of people have the same birthday, since then $\mathbb{P}[A]$ is larger than 50%. For $n = 60$ people, $\mathbb{P}[A]$ is over 99%.

3.7 The Monty Hall Problem

In an (in)famous 1970s game show hosted by Monty Hall, a contestant was shown three doors; behind one of the doors was a valuable prize (a car), and behind the other two were goats. The contestant picks a door (but does not 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/her current door, or switching to the other unopened one. The contestant wins the prize if and only if their chosen door is the correct one. The question is: Does the contestant have a better chance of winning if he/she 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/her 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. (Actually our calculation will have the same result no matter how Carol picks the door.)

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

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

[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

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

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 W we are interested in is the event that the contestant wins. Which sample points (i, j, k) are in W? Well, since the contestant is switching doors, their 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 W 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 $\mathbb{P}[W] = \frac{6}{9} = \frac{2}{3}$. 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 their probability of winning is $\frac{1}{3}$. (In this case, the contestant is really just picking a single random door.) So by switching, the contestant actually improves their odds by a huge amount!

4 Summary

The Monty Hall example well illustrates the importance of doing probability calculations systematically, rather than "intuitively." Recall the key steps in all our calculations:

- What is the sample space (i.e., the experiment and its set of possible outcomes)?
- What is the probability of each outcome (sample point)?
- What is the event 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 contained in 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 claim that the switching strategy in the Monty Hall problem does not improve the odds.