CS 70 Discrete Mathematics for CS
Fall 2004 Rao

Lecture 14

Introduction to Probability

The next several lectures will be concerned with probability theory. We will aim to make sense of statements such as the following:

1. "There is a 30% chance of a magnitude 8 earthquake in Northern California before 2030."
2. "The average time between system failures is about three days."
3. "The chance of getting a flush in a five-card poker hand is about 2 in 1000."
4. "In this load-balancing scheme, the probability that any processor has to deal with more than twelve requests for service is negligible."

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

Probability spaces

Every probability space is based on a random experiment, such as rolling a die, shuffling a deck of cards, picking a number, assigning jobs to processors, running a system etc. Rather than attempt to define "experiment" directly, we shall define it by its set of possible outcomes, which we call a "sample space." Note that all outcomes must be disjoint, and they must cover all possibilities.

Definition 14.1 (sample space): The sample space of an experiment is the set of all possible outcomes. An outcome is often called a sample point or atomic event.

Definition 14.2 (probability space): A probability space is a sample space Ω, together with a probability $\Pr[\omega]$ for each sample point ω, 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.

[Strictly speaking, what we have defined above is a restricted set of probability spaces known as *discrete* spaces: this means that the set of sample points is either finite or countably infinite (such as the natural numbers, or the integers, or the rationals, but *not* the real numbers). Later, we will talk a little about continuous sample spaces, but for now we assume everything is discrete.]

Here are some examples of (discrete) probability spaces:

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 ith toss. Thus Ω 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 $\frac{1}{2^n}$.

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. 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 (one red, one blue). Then $\Omega = \{(i,j) : 1 \leq i,j \leq 6\}$. Each of the 36 outcomes has equal probability, $\frac{1}{36}$.

5. Roll two indistinguishable dice (i.e., both blue). Here $\Omega = \{(i,j) : 1 \leq i \leq j \leq 6\}$. There are 21 outcomes (why?); those of the form (i,i) have probability $\frac{1}{36}$, and the rest have probability $\frac{2}{36} = \frac{1}{18}$. [How did we derive these probabilities?] When the dice are indistinguishable, we are forced to use this sample space. We might also use it, for example, when modeling a game of backgammon, even when the two dice *are* distinguishable: for we are only interested in the two numbers that come up, not in their order.

6. **Card shuffling.** Shuffle a deck of cards. Here Ω 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.]

7. **Poker hands.** Shuffle a deck of cards, and then deal a poker hand. Here Ω 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).[1] Recall that $\binom{n}{k} = \frac{n!}{k!(n-k)!}$. Consequently, $\binom{52}{5} = \frac{52!}{5!47!} = \frac{52\times51\times50\times49\times48}{5\times4\times3\times2\times1} = 2598960$. Such experiments, in which some number of *distinct* elements are drawn from a set of size n, are usually known as "sampling without replacement."

8. **Prize draws.** n contestants enter a prize draw; three distinct winners are picked at random, in order (first, second, third). Here Ω consists of all triples of the form (i,j,k), where $i,j,k \in \{1,\ldots,n\}$ and i,j,k are distinct. Each sample point has equal probability. The number of sample points, $|\Omega|$, is equal to the number of ways of picking three elements, in order, out of a set of size n: this was denoted $P(n,3)$ in Lecture 9, where we also saw that $P(n,k) = \frac{n!}{(n-k)!}$. So we have $|\Omega| = \frac{n!}{(n-3)!} = n(n-1)(n-2)$. This is another example of sampling without replacement, except that in this case the order of the chosen elements is significant.

[1] The notation $\binom{n}{k}$ is pronounced "n choose k". Some writers use the alternative notation $C(n,k)$, which is entirely equivalent. In either case, it is usually pronounced "n choose k."

9. **Balls and bins.** Throw 20 (distinguishable) balls into 10 (distinguishable) bins, so that each ball is equally likely to land in any bin, regardless of what happens to the other balls. Here $\Omega = \{(b_1, b_2, \ldots, b_{20}) : 1 \leq b_i \leq 10\}$; 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 $n = 2$.] This is an example of "sampling *with* replacement," since the choices made by the balls are not required to be distinct. 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.

10. **Balls and bins with indistinguishable balls.** Suppose we throw m balls into n bins, but that now the balls are indistinguishable from one another: this means that, after throwing the balls, we see only the number of balls that landed in each bin, but not which balls landed where. Thus each sample point is just an n-tuple $(m_1, m_2, \ldots, m_n)$, in which m_i is the number of balls that land in bin i: i.e., Ω consists of all such n-tuples in which the m_i are non-negative integers that sum to m. How many sample points are there? This is equal to the number of ways of partitioning m indistinguishable items into n groups. We can think of this as follows: write down m dots in a line (representing the items); then take $n-1$ vertical lines and interleave them any way you like with the dots. Any such arrangement of dots and lines represents a unique partition as follows: reading along the line from left to right, the number of items in the first group, m_1, is the number of dots to the left of the first line. The number of items in the second group, m_2, is the number of dots between the first and second lines, and so on. It should be clear that each partition is represented by one and only one dot-line arrangement, so the number of partitions is equal to the number of arrangements. But the number of dot-line arrangements is easy to compute: it is $\binom{m+n-1}{n-1}$, because we just have to choose the positions for the $n-1$ lines out of the total number of available positions $m+n-1$. (Think about it!)

    In this sample space, the probabilities of sample points are *not* all equal. For example, if we have two balls and two bins, then the arrangement $(2,0)$ (in which both balls fall into bin 1) can happen in only one way, whereas $(1,1)$ can happen in two different ways (switching the two balls over gives us the same sample point). In general, how many ways can we get the arrangement $(m_1, m_2, \ldots, m_n)$? Well, by generalizing our proof in Lecture 9 for the number of ways of choosing k things from n, the answer is $\frac{m!}{m_1! \times m_2! \times \cdots \times m_n!}$. So this sample point ω should have probability $\Pr[\omega] = \frac{1}{n^m} \times \frac{m!}{m_1! \times m_2! \times \cdots \times m_n!}$.

11. **The Monty Hall Problem.** In the (in)famous 1960s game show "Let's Make a Deal," 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), and Monty opens one of the other two doors, revealing a goat. 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?

    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 Monty. Note that some triples are not possible: e.g., $(1, 2, 1)$ is not, because Monty 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 Monty to open), Monty 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 Monty opening door 3 (with probability 1, because he has no choice). So

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

[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

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

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 $(6 \times \frac{1}{9}) + (6 \times \frac{1}{18}) = 1$.

Events

In the Monty Hall problem, we are interested in the probability that the constestant wins the prize. This is itself not a single outcome (the contestant can win in several different ways), but a *set* of outcomes. This leads us to:

Definition 14.3 (event): An event A in a sample space Ω is any subset $A \subseteq \Omega$.

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

Definition 14.4 (probability of an event): For any event $A \subseteq \Omega$, we define the probability of A to be

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

Let's look at some examples; the number of the example refers to the probability space in our previous list.

1. **Fair coin.** Let A be the event "the coin comes up Heads." Then $\Pr[A] = \frac{1}{2}$.

2. **Three fair coins.** Let A be the event that all three coin tosses are the same. Then $\Pr[A] = \Pr[HHH] + \Pr[TTT] = \frac{1}{8} + \frac{1}{8} = \frac{1}{4}$.

3. **Biased coins.** 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 $\binom{n}{r} p^r (1-p)^{n-r}$.

4. **Dice.** Let A be the event 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 uniform, i.e., all the sample points have the *same* probability (which must be $\frac{1}{|\Omega|}$, where $|\Omega|$ denotes the size of Ω). 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!

6. **Card shuffling.** Let A be the event that the top card is an ace. Then by the above remarks

$$\Pr[A] = \frac{\text{\# of permutations with an ace on top}}{52!}.$$

How many permutations have an ace on top? Well, there are four choices for the ace; and once we have chosen it and put it on top, there are exactly 51! ways to arrange the remaining 51 cards. So the number of such permutations is $4 \times 51!$. Thus $\Pr[A] = \frac{4\times 51!}{52!} = \frac{4}{52} = \frac{1}{13}$.

Let B be the event that the top two cards are from the same suit. How many such permutations are there? Well, there are four possible common suits for the top two cards, and once the suit is chosen there are 13 ways to choose the first card and 12 ways to choose the second. The remaining 50 cards can be arranged in 50! ways. Thus we have

$$\Pr[B] = \frac{4\times 13\times 12\times 50!}{52!} = \frac{12}{51}.$$

7. **Poker hands.** What is the probability that our poker hand is a flush? [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 $4 \times \binom{13}{5}$. So we have

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

8. **Prize draws.** I am one of the n contestants. Let A be the event that I win one of the three prizes. How many sample points lie in A? Well, there are three prizes I could win; each of these leaves two prize slots which have to be filled by choosing two contestants (in order) from the remaining $n-1$; and there are $P(n-1,2) = (n-1)(n-2)$ ways to do this. So $|A| = 3(n-1)(n-2)$, and $\Pr[A] = \frac{|A|}{|\Omega|} = \frac{3(n-1)(n-2)}{n(n-1)(n-2)} = \frac{3}{n}$.

9. **Balls and bins.** 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

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

11. **Monty Hall.** Finally, 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 Monty 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 $\Pr[A] = \frac{6}{9} = \frac{2}{3}$. I.e., 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.