CS 30: Discrete Math in CS (Winter 2019): Lecture 21

Date: 13th February, 2019 (Wednesday)
Topic: Probability: Basics
Disclaimer: These notes have not gone through scrutiny and in all probability contain errors.
Please discuss in Piazza/email errors to deeparnab@dartmouth.edu

---

1. **Experiments and Outcomes : Sample Spaces.**

   Every time you hear a question beginning with, "What are the chances that ... ", an *experiment* with an (yet) unknown *outcome* has taken place. Let us consider a few examples:

   - What is the chance that a tossed coin lands heads?
   - What is the chance we get a score of 7 with two rolled dice?
   - What is the chance that two people in a party share a birthday?

   In each of these questions above, there are experiments involved. We toss a coin. We roll two dice. We note down people's birthdays. Each of these experiments have outcomes involved. In the first one, the outcomes are heads or tails. In the second one, the outcome can be noted in two ways; we can either note down the sum and observe that the outcomes are a number between 2 and 12, or we can just store down the two numbers we see as a tuple, like $(3, 4)$. In the third one, the outcome is a list of birthdays of the people in the party.

   The *first step* in answering in any question of the form above, we must be sure what the *experiment* is, and what the possible *outcomes* are. This forms the ***sample space*** of the question at hand. ✍

   > **Exercise:** *What are the experiments and possible outcomes for the following questions (google the definitions if you don't know them):*
   >
   > - *What is the chance of getting a royal flush in a five hand poker draw?*
   > - *What is the chance of getting four aces in a bridge hand?*
   > - *What is the chance of seeing a run of* 5 *heads when you toss a coin* 100 *times?*

2. **Events.**

   Ok. So you have figured out the *sample space* of the question at hand, and let us call Ω the set of all possible outcomes. So in the first question $\Omega = \{\texttt{heads}, \texttt{tails}\}$, for the second question Ω could be the set $\{2, 3, \ldots, 11, 12\}$, or it could be the set $\{(1, 1), (1, 2), \ldots, (6, 5), (6, 6)\}$. For the third question, Ω is *all* possible lists of birthdays. How big is it?

   We go back to the question and figure out the "set of outcomes we are interested in". Sometimes it is just one outcome; in the first question, we are only interested in the outcome `heads`. In the second question, if we thought of our sample space as $\{2, 3, \ldots, 11, 12\}$, then we are just interested in one outcome $\{7\}$. However, if we thought of our sample space as $\{(1, 1), (1, 2), \ldots, (6, 5), (6, 6)\}$, then we are actually interested in multiple outcomes such as $(3, 4)$ or $(6, 1)$, and so on. Similarly, in the third event, there are multiple outcomes which are of interest; any list which has *at least* two repeating birthdays is of interest to us.

In general, a question of the form "what are the chances" asks the question of an *event*. Formally, an event $\mathcal{E} \subseteq \Omega$ is the set of outcomes which lead to the event in consideration.

3. **Randomness and Probabilities.**

   Right. Now we have the sample space Ω figured out. We have also figured out the relevant outcomes $\mathcal{E} \subseteq \Omega$ which we are interested in. Now comes the time to assign *likelihoods* to all the outcomes in Ω so that we can reasonably answer the question at hand. Formally, if we let Ω denote the set of possible outcomes, then the likelihood of any outcome $o \in \Omega$ is a "score" called the *probability* $\mathbf{Pr}[o]$ which is a *real number* between 0 and 1. Apart from this, the only constraint on $\mathbf{Pr}$ is that $\sum_{o \in \Omega} \mathbf{Pr}[o] = 1$.

   This function $\mathbf{Pr}$ which assigns a non-negative real value to each outcome is called the *probability distribution*. If $\mathbf{Pr}[o] = 1/|\Omega|$ for all $o \in \Omega$, then the distribution is called the *uniform distribution over the sample space*.

   How does one assign these scores? Ultimately, at some level these are determined by the *modeling assumptions* of the experiment at hand. However, what math allows us to do is to make as few and as *elementary* assumptions as possible. Let's discuss this by considering the questions again from the previous bullet point.

   - *What is the chance that a tossed coin lands heads?* The sample space is $\Omega = \{\texttt{heads}, \texttt{tails}\}$. What is $\mathbf{Pr}[\texttt{heads}]$? Well, if we assume that the coin was fair, and the toss was fair, then it is *reasonable* to assume that $\mathbf{Pr}[\texttt{heads}] = \mathbf{Pr}[\texttt{tails}]$, and therefore since they sum to 1, each of these should be $1/2$. This is *one* model. Another model could be that perhaps the coin is not fair, but in that case the model must specify what the probability of getting a heads is.
   - *What is the chance we get a score of 7 with two rolled dice?* Let's start with the first sample space here is $\Omega = \{2, 3, 4, \ldots, 11, 12\}$. What is $\mathbf{Pr}[2]$? Is it *reasonable* to assume now that all the numbers in Ω should have the same likelihood? A little thought tells us no. What would be reasonable to assume is that each of the two die is fair, that is, each die returns a answer in $\{1, 2, 3, 4, 5, 6\}$ equally likely. That is an *elementary* assumption and a more reasonable assumption. Indeed, this assumption will let us argue that if we took $\Omega = \{(1,1), (1,2), \ldots, (6,5), (6,6)\}$, then the probability would indeed be uniform over this sample space.
   - *What is the chance that two people in a party share a birthday?* This is slightly trickier. The sample space Ω, remember, is a list of dates. Are all lists equally likely? Is that a reasonable assumption? Depends. In truth, it doesn't seem to be so. Data shows that more people are born in September than in February. So the lists should have more September than Februaries. But for the *purpose of a first cut*, it doesn't seem to be too bad an assumption. And we will make this assumption to answer this question.

   Ok, so where are we? When faced with a question we figure out the sample space Ω. We make our modeling assumptions, make them as simple and as reasonable as possible. Sometimes these assumptions immediately lead to the $\mathbf{Pr}[o]$ for each $o \in \Omega$ (like the coin question above, like the birthday question above). Sometimes not (for the dice question above).

However, if we figure out the $\mathbf{Pr}[o]$ for every $o \in \Omega$, then we can answer the question using

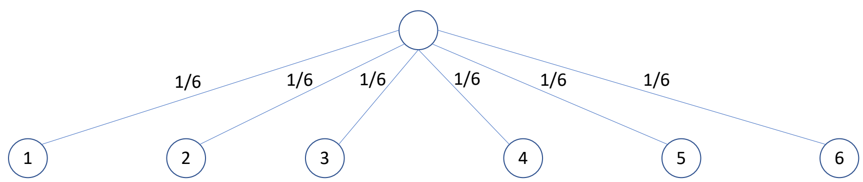
$$\mathbf{Pr}[\mathcal{E}] = \sum_{o \in \mathcal{E}} \mathbf{Pr}[o]$$

And indeed, if the distribution is uniform, then $\mathbf{Pr}[o] = \frac{1}{|\Omega|}$ then, $\mathbf{Pr}[\mathcal{E}] = \frac{|\mathcal{E}|}{|\Omega|}$.

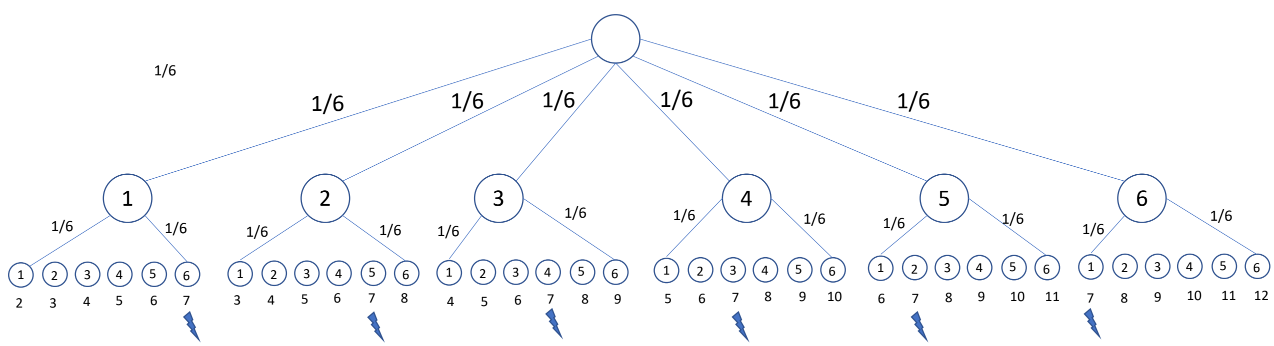
4. **Figuring out Outcome Probabilities and the Probability of the Event.**

   Let's figure out the outcome probabilities for the dice question. We do so by drawing the "tree diagram". This is very useful technique when there are "multiple" elementary assumptions which lead to the outcomes.

   Back to the dice question. We first roll the first dice, and draw the situations as a tree below. The outcome is written in the circle. The probabilities are written on the edge. Now, we roll



   the second dice. We write the outcome of the second dice in the circles in the second level. The sums are written below. The "lightning" ones are that we are interested in.

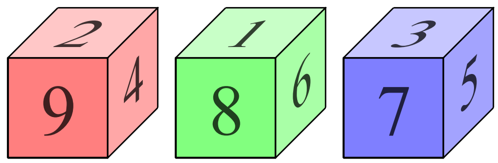


   The probability of each outcome in the second level of the tree is $\frac{1}{6} \cdot \frac{1}{6} = \frac{1}{36}$. Since there are six of these outcomes which lead to the sum of 7, the probability of the event that we see a 7 is indeed $6/36 = 1/6$.

Exercise: *What is the probability that the sum of two rolled dice leads to a prime number?*

5. **Intransitive Dice.**

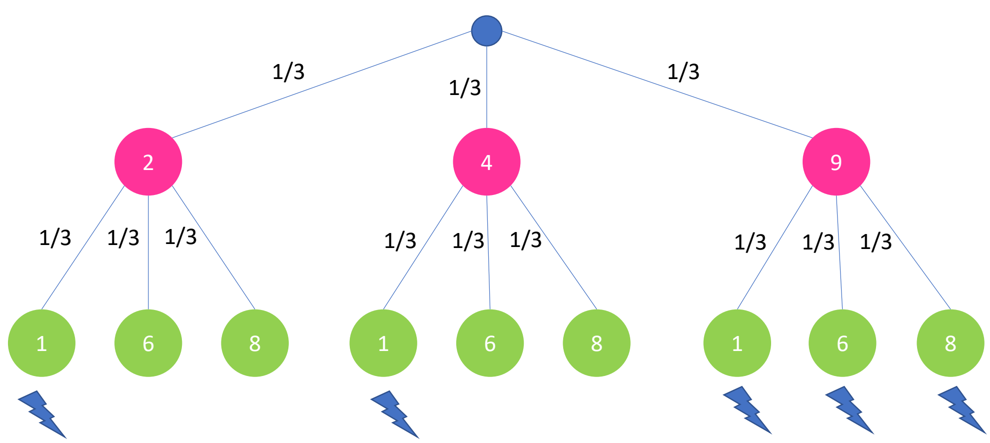
   Consider the following whacky dice. Each dice has the same number on the two opposite sides.



Let's call the red dice R, the green dice G, and the blue dice B. Fix two of these: say R and G. We roll both of them. We are interested in the event that the red die gets a bigger number than the green die. ✍

Exercise: *What is the probability that the red die gets a bigger number than the green die when both of them are rolled once?*

Once again, tree diagrams are the perfect procedure to answer this question. The following tree shows the picture where first we rolls the red die, and then the green die. The various outcomes are put in the circles. The "lightning" denotes the event that the red die gets a bigger number than the green die. Therefore, the probability that the red die



gets a bigger number than the green die is $5/9$, since each outcome in the bottom layer occurs with probability $1/9$, and the "lightning strikes 5 times", that is, the event $\mathcal{E} = \{\text{Red Die has larger number than Blue Die}\}$ occurs 5 times out of 9.

Therefore, you see that the Red die is a better die than Green die. We denote this as $R \succ G$ to show R "beats" G. ✍

Exercise: *Drawing the tree diagrams, what is the probability that the green die beats the blue die with one roll? What is the probability that the blue die beats the red die? Do you see something counter-intuitive?*

6. **The Monty Hall Problem.** Here is a question you may or may not have heard before:

> *Suppose you're on a game show, and you're given the choice of three doors: Behind one door is a car; behind the others, goats. You pick a door, say No. 1, and the host, who knows what's behind the doors, opens another door, say No. 3, which has a goat. He then says to you, "Do you want to pick door No. 2?" Is it to your advantage to switch your choice?*

At first glance, it may not seem like a question in probability at all. And indeed, it is ill-formulated. To formulate this well, we need to define the experiments, define the outcomes, clearly state the modeling assumptions, calculate the probabilities based on this, and then answer.

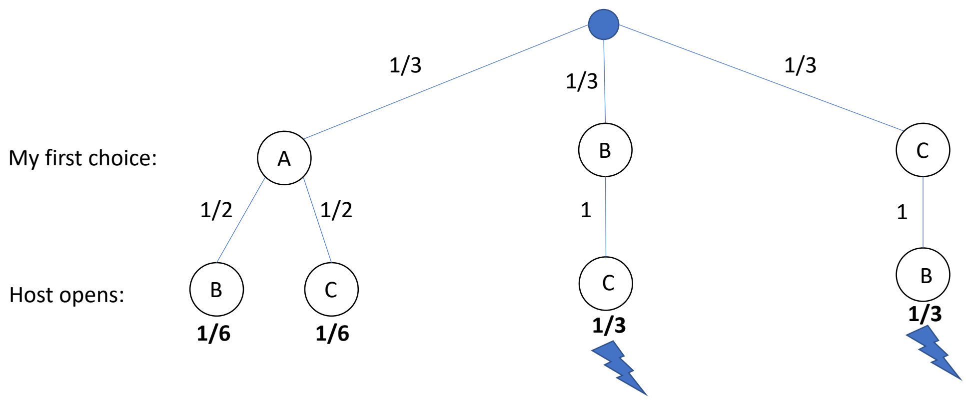
So here are the assumptions which I would make depending on my reading:

- The first door I pick is uniformly at random between the three doors. I don't know better.
- The host who does know the answer chooses a door to open. He is constrained to open a door which has a goat. However, my belief is that if there are two door which has goats, then he would open one of the two also uniformly at random. Once again, my rationale is that the question doesn't specify it, and so I assume he doesn't know better.

Once we have fixed this, then we can ask the question "What is the probability that the 'third door' has the car?" To answer this, again, we can use tree diagrams.

Before we begin, let us *rename* the doors such that the car is behind Door 1, and Door 2 and 3 contains goats. This is without loss of generality – my choice and the host's choice, by my assumption, doesn't depend on the numbers. The tree diagram is below. The first layer shows the random choices for the first door. The second layer shows the random choices of the door opened by the host. Note, that it is random only in the first branch; if the door I choose in the first try has a goat, then the host has to open the third door (the one not containing the car, and the one I chose). The "lightning" indicates the outcomes in which switching leads me to the car. The numbers in bold indicates the probability of that outcome.

As you can see from the tree-diagram, the probability switching helps is $2/3$.



7. **Notation.**

With events, we often mix-and-match notation from Boolean Logic and Sets.

- Given an event $\mathcal{E}$, the negation event $\neg\mathcal{E}$ is used to denote the event that $\mathcal{E}$ doesn't take place. That is, it is simply the subset $\neg\mathcal{E} = \Omega \setminus \mathcal{E}$. Sometimes, $\neg\mathcal{E}$ is denoted as $\overline{\mathcal{E}}$.

$$\mathbf{Pr}[\mathcal{E}] + \mathbf{Pr}[\neg\mathcal{E}] = 1$$

- Given two events $\mathcal{E}$ and $\mathcal{F}$, the notation $\mathcal{E} \cup \mathcal{F}$ is precisely the union of the subsets in the sample space. $\mathbf{Pr}[\mathcal{E} \cup \mathcal{F}]$ captures the likelihood that at least one of the events takes place.
- Given two events $\mathcal{E}$ and $\mathcal{F}$, the notation $\mathcal{E} \cap \mathcal{F}$ is precisely the intersection of the subsets in the sample space. $\mathbf{Pr}[\mathcal{E} \cap \mathcal{F}]$ captures the likelihood that at least one of the events takes place.
- Two events $\mathcal{E}$ and $\mathcal{F}$ are *disjoint* or *exclusive* if $\mathcal{E} \cap \mathcal{F} = \emptyset$. That is, they both can't occur simultaneously. A collection of events $\mathcal{E}_1, \mathcal{E}_2, \ldots, \mathcal{E}_k$ are *mutually exclusive* if $\mathcal{E}_i \cap \mathcal{E}_j = \emptyset$ for $i \neq j$.
- For mutually exclusive events,

$$\mathbf{Pr}[\mathcal{E}_1 \cup \mathcal{E}_2 \cup \cdots \mathcal{E}_k] = \sum_{i=1}^{k} \mathbf{Pr}[\mathcal{E}_i]$$

- The Inclusion-Exclusion formula (for two events, aka Baby version) tells us

$$\mathbf{Pr}[\mathcal{E} \cup \mathcal{F}] = \mathbf{Pr}[\mathcal{E}] + \mathbf{Pr}[\mathcal{F}] - \mathbf{Pr}[\mathcal{E} \cap \mathcal{F}]$$

  Do you see why? It is *exactly* the baby-version of inclusion-exclusion if $\mathbf{Pr}$ is a *uniform distribution*. Indeed, if this is the case then $\mathbf{Pr}[\mathcal{E} \cup \mathcal{F}] = \frac{|\mathcal{E} \cup \mathcal{F}|}{|\Omega|}$, and the proof follows by applying baby IE. What if it is not uniform?

Exercise: *Prove the above even when* $\mathbf{Pr}$ *is* not *a uniform distribution.*