CS 30: Discrete Math in CS (Winter 2019): Lecture 22

Date: 15th February, 2019 (Friday)
Topic: Probability: Conditional Probability and Independence
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. **Some Recap.**

   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 both 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.*

2. **Conditional Probability.**

   Last time, we looked at (semi) formal definitions of sample spaces, events, probability of outcomes, and probability of events. We now look at a very important concept of *conditional probability*. In plain English, these are trying to answer questions of the form

   What are the chances of "blah" happening, if we know that "blooh" has already occured?

Concrete examples:

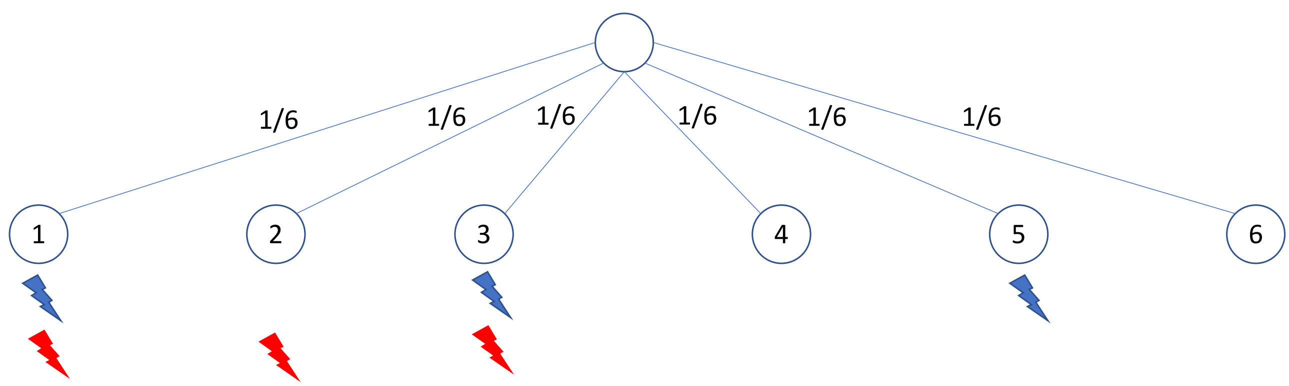
- What is the probability that a roll of fair die lies in the set $\{1, 2, 3\}$ given that the roll is an odd number?
- What is the probability that a roll of two fair dice sums to 6 given that the sum is an even number?

In both these questions above, there are *two* events of interest. For example, in the first example, one event is $\mathcal{A}$ which occurs when the roll of the fair die lies in the set $\{1, 2, 3\}$ (this is really the event we are interested in). But there is also another event, let's call it $\mathcal{B}$, which occurs when the roll of the fair die is an odd number. The first question is asking, what is the probability that $\mathcal{A}$ occurs given that $\mathcal{B}$ has already occurred.

This probability is *different* then just $\mathbf{Pr}[\mathcal{A}]$ or just $\mathbf{Pr}[\mathcal{B}]$. It is called the *conditional probability* of event $\mathcal{A}$ occuring *given* that $\mathcal{B}$ has already occured. And it is denoted as

$$\mathbf{Pr}[\mathcal{A} \mid \mathcal{B}]$$

We will derive the formula for the above, but before that, let's solve the question one above using a tree diagram. Below is the tree diagram for a single dice throw. The "blue lightnings" (the ones on top) indicate the outcomes which lead to the even $\mathcal{B}$, that is, the die comes out odd. The "red lightning" (the one on bottom) indicates the outcome $\mathcal{A}$ which we are interested in.



When calculating the conditional probability, we are guaranteed that the "blue lightning" has struck, and among all the outcomes in which the blue lightning strikes, what is the likelihood that the red lightning strikes *as well*. Therefore, when trying to figure out $\mathbf{Pr}[\mathcal{A} \mid \mathcal{B}]$, the *sample space has changed*! It is not Ω any more, but rather it is $\mathcal{B}$.

Since Ω has changed, $\mathbf{Pr}[\cdot]$ must change too. How should it change? Here is another assumption we make: given only the promise that $\mathcal{B}$ occurs, the *relative likelihood* of two different outcomes in $\mathcal{B}$ shouldn't change. Thus, the new probability distribution, let's call it $\mathbf{Pr}'$, over $\mathcal{B}$, should be a scaled version of the old distribution only over the new sample space

$\mathcal{B}$. That is, $\mathbf{Pr}'[o] = c \cdot \mathbf{Pr}[o]$ for all $o \in \mathcal{B}$ where c is some number which guarantees that $\sum_{o \in \mathcal{B}} \mathbf{Pr}'[o] = 1$. This implies $c = \frac{1}{\mathbf{Pr}[\mathcal{B}]}$. Now, in this new probability distribution $\mathbf{Pr}'$, we are interested in the outcomes when $\mathcal{A}$ occurs. Again note, we are not interested in the outcomes when $\mathcal{A}$ occurs when $\mathcal{B}$ *doesn't occur*. We are only interested in the outcomes in $\mathcal{A} \cap \mathcal{B}$ (when both the red and blue lightning strike). Thus, we get $\mathbf{Pr}[\mathcal{A} \mid \mathcal{B}] = \mathbf{Pr}'[\mathcal{A} \cap \mathcal{B}]$. Using the fact that $\mathbf{Pr}'[o] = \mathbf{Pr}[o]/\mathbf{Pr}[\mathcal{B}]$ for all $o \in \mathcal{B}$, we get the following formula for conditional probability. Tattoo this in your head.

$$\mathbf{Pr}[\mathcal{A} \mid \mathcal{B}] := \frac{\mathbf{Pr}[\mathcal{A} \cap \mathcal{B}]}{\mathbf{Pr}[\mathcal{B}]} \qquad \text{(Cond Prob)}$$

Coming back to the dice problem number 1, $\mathbf{Pr}[\mathcal{B}] = 1/2$ and $\mathbf{Pr}[\mathcal{A} \cap \mathcal{B}] = 2/6$, thus, the probability that the dice gives a number in $\{1, 2, 3\}$ when given that the dice gives an odd number is $2/3$.

✍

Exercise: *Solve the second dice problem: what is the probability that a roll of two fair dice sums to 6 given that the sum is an even number? Use both: the method of conditional probabilities, and the tree diagram from last time. Are your answers the same?*

✍

Exercise: *I roll two dice. $\mathcal{A}$ be the event that the first dice is odd. $\mathcal{E}$ is the event that the sum of the two dice is odd. What is $\mathbf{Pr}[\mathcal{A} \mid \mathcal{E}]$?*

3. **Chain Rule.**

A simple but important consequence of the definition of conditional probability is the *chain rule*.

Theorem 1. For any two events $\mathcal{A}$ and $\mathcal{B}$, we have $\mathbf{Pr}[\mathcal{A} \cap \mathcal{B}] = \mathbf{Pr}[\mathcal{B}] \cdot \mathbf{Pr}[\mathcal{A} \mid \mathcal{B}]$. More generally, for any collection of events $\mathcal{A}_1, \mathcal{A}_2, \cdots, \mathcal{A}_k$, we have

$$\mathbf{Pr}[\mathcal{A}_1 \cap \mathcal{A}_2 \cap \cdots \mathcal{A}_k] = \mathbf{Pr}[\mathcal{A}_1] \cdot \mathbf{Pr}[\mathcal{A}_2 \mid \mathcal{A}_1] \cdot \mathbf{Pr}[\mathcal{A}_3 \mid \mathcal{A}_1 \cap \mathcal{A}_2] \cdots \mathbf{Pr}[\mathcal{A}_k \mid \mathcal{A}_1, \mathcal{A}_2, \ldots, \mathcal{A}_{k-1}]$$

Applications

- Here's an example (from the textbook) showing how this is useful: *what's the probability that 5 randomly drawn cards from a standard deck are all hearts?*

  Think of drawing these cards one by one from the deck to your hand. Let $\mathcal{A}_i$, for $i = 1, 2, \ldots, 5$ be the event that the ith card is a heart. We need to figure out $\mathbf{Pr}[\mathcal{A}_1 \cap \mathcal{A}_2 \cdots \cap \mathcal{A}_5]$.

  Note:

  – $\mathbf{Pr}[\mathcal{A}_1] = \frac{13}{52}$; there are 13 hearts to begin with, and 52 cards in all.
  – $\mathbf{Pr}[\mathcal{A}_2 \mid \mathcal{A}_1] = \frac{12}{51}$. Why? Given that $\mathcal{A}_1$ has occurred, the deck now is one heart missing. Thus, there are 51 cards in all and only 12 of them are hearts.
  – Similarly continuing, we get $\mathbf{Pr}[\mathcal{A}_3 \mid \mathcal{A}_1, \mathcal{A}_2] = \frac{11}{50}$; $\mathbf{Pr}[\mathcal{A}_4 \mid \mathcal{A}_1, \mathcal{A}_2, \mathcal{A}_3] = \frac{10}{49}$; $\mathbf{Pr}[\mathcal{A}_5 \mid \mathcal{A}_1, \mathcal{A}_2, \mathcal{A}_3, \mathcal{A}_4] = \frac{9}{48}$.

– Thus, $\mathbf{Pr}[\mathcal{A}_1 \cap \cdots \mathcal{A}_5] = \frac{13}{52} \cdot \frac{12}{51} \cdot \frac{11}{50} \cdot \frac{10}{49} \cdot \frac{9}{48}$

•

✍

Exercise: *Suppose we take a random ordering of the elements $(1, 2, 3, \ldots, n)$. What is the probability that 1 is in the first place, and 2 is in the second place, 3 is in the third place, 4 is in the fourth place, and 5 is in the fifth place of this random ordering?*

4. **The Law of Total Probability.**

Sometimes conditioning *helps* in figuring out probability of events. That is, suppose we are interested in finding the probability of event $\mathcal{A}$. Sometimes this is easier to do if we already know whether some event $\mathcal{B}$ has taken place or not. Then, we can use the following formula to figure out the probability of $\mathcal{A}$.

Theorem 2. For any two events $\mathcal{A}$ and $\mathcal{B}$,

$$\mathbf{Pr}[\mathcal{A}] \quad = \quad \mathbf{Pr}[\mathcal{A} \mid \mathcal{B}] \cdot \mathbf{Pr}[\mathcal{B}] \quad + \quad \mathbf{Pr}[\mathcal{A} \mid \neg\mathcal{B}] \cdot \mathbf{Pr}[\neg\mathcal{B}]$$

Proof. The proof follows by noticing that the event (subset) $\mathcal{A}$ can be partitioned into two *disjoint* subsets as follows:

$$\mathcal{A} = (\mathcal{A} \cap \mathcal{B}) \cup (\mathcal{A} \cap \neg\mathcal{B})$$

Convince yourself of this fact.

Thus, $\mathbf{Pr}[\mathcal{A}] = \mathbf{Pr}[\mathcal{A} \cap \mathcal{B}] + \mathbf{Pr}[\mathcal{A} \cap \neg\mathcal{B}]$. And the theorem follows from the formula for conditional probability. □

In fact, there are two successive generalizations which at times are useful.

Theorem 3. Let $\mathcal{B}_1, \mathcal{B}_2, \ldots, \mathcal{B}_k$ be *mutually exclusive* events (that is pairwise disjoint) such that $\sum_{i=1}^{k} \mathbf{Pr}[\mathcal{B}_i] = 1$. Then,

$$\mathbf{Pr}[\mathcal{A}] = \sum_{i=1}^{k} \mathbf{Pr}[\mathcal{A} \mid \mathcal{B}_i] \cdot \mathbf{Pr}[\mathcal{B}_i]$$

Theorem 4. Let $\mathcal{B}_1, \mathcal{B}_2, \ldots, \mathcal{B}_k$ be *mutually exclusive* events (that is pairwise disjoint) with $\mathcal{B}_1 \cup \mathcal{B}_2 \cup \cdots \cup \mathcal{B}_k = \mathcal{B}$, Then,

$$\mathbf{Pr}[\mathcal{A} \mid \mathcal{B}] = \sum_{i=1}^{k} \mathbf{Pr}[\mathcal{A} \mid \mathcal{B}_i] \cdot \mathbf{Pr}[\mathcal{B}_i \mid \mathcal{B}]$$

✍

Exercise: *Prove the above two theorems. Exactly the same idea as the proof of the previous theorem.*

Applications:

- *In a bag there are two coins. One is a fair coin which, when tossed, lands heads with probability* 0.5*. The other, however, is a biased coin which, when tossed, lands heads with probability* 0.75*. You pick one of the two coins at random. What is the probability you see heads?*

  You could do this with a tree diagram, but we can also do with the above law of total probability (it is the same thing!). Let $\mathcal{A}$ be the event that we see heads; we are interested in $\mathbf{Pr}[\mathcal{A}]$. Let $\mathcal{B}$ be the event we pick a fair coin; so $\neg\mathcal{B}$ is the event we pick the biased coin.

  We know, by the problem definition, $\mathbf{Pr}[\mathcal{A} \mid \mathcal{B}] = 0.5$ and $\mathbf{Pr}[\mathcal{A} \mid \neg\mathcal{B}] = 0.75$. Furthermore, since we pick one of the two coins at random, we gt $\mathbf{Pr}[\mathcal{B}] = 0.5$. Therefore, by the law of total probability,

  $$\mathbf{Pr}[\mathcal{A}] = (0.5) \cdot (0.5) + (0.75) \cdot (0.5) = 0.625$$

5. **Independent and Dependent Events.**

   In the example above, the probability that a roll of a fair die is 3 if nothing more is told (the answer is $1/6$) is *different* from the probability that a roll of a fair die is 3 given that the roll is an odd number (the answer is $1/3$). Thus, the event $\mathcal{B}$, that the roll was odd, told us something about the event $\mathcal{A}$ whether the roll was 3. $\mathcal{B}$ had some *dependence* on $\mathcal{A}$.

   But many times two events may not show such dependence. For example, consider having two dice. Let $\mathcal{A}$ be the event that the first dice rolls a 3. Let $\mathcal{B}$ be the event that the second dice rolls an odd number. Would $\mathbf{Pr}[\mathcal{A}]$ and $\mathbf{Pr}[\mathcal{A} \mid \mathcal{B}]$ be different? You may feel of course not – what does the roll of the second die have to do with the roll of the first die? And you would be correct. Nevertheless, let's just calculate $\mathbf{Pr}[\mathcal{A} \mid \mathcal{B}]$ in this example.

   $$\mathbf{Pr}[\mathcal{A} \mid \mathcal{B}] = \frac{\mathbf{Pr}[\mathcal{A} \cap \mathcal{B}]}{\mathbf{Pr}[\mathcal{B}]} = \frac{\frac{3}{36}}{\frac{3}{6}} = \frac{1}{6} = \mathbf{Pr}[\mathcal{A}]$$

   where the numerator can be found by drawing the tree diagram as last time. Indeed, the only outcomes which lead to $\mathcal{A} \cap \mathcal{B}$ are $\{(3,1),(3,3),(3,5)\}$.

   This brings us to a very, very important definition.

   Remark: *Given a random experiment, two events* $\mathcal{A}$ *and* $\mathcal{B}$ *are* **independent** *if and only if* $\mathbf{Pr}[\mathcal{A} \mid \mathcal{B}] = \mathbf{Pr}[\mathcal{A}]$*. Equivalently,*

   $$\mathbf{Pr}[\mathcal{A} \cap \mathcal{B}] = \mathbf{Pr}[\mathcal{A}] \cdot \mathbf{Pr}[\mathcal{B}]$$

   Often times, when the outcomes of the two events in consideration are "generated" using different "sources of uncertainty" (eg, the two dice in the previous example), then these are independent events. Here are some examples of independent events. Confirm this by figuring out $\mathbf{Pr}[\mathcal{A} \cap \mathcal{B}]$, $\mathbf{Pr}[\mathcal{A}]$, and $\mathbf{Pr}[\mathcal{B}]$.

- Two coins are tossed. $\mathcal{A}$ is the event the first lands heads, $\mathcal{B}$ is the event that the second lands tails.
- An n-length bit string is picked at random from all n-length bit strings. $\mathcal{A}$ is the event that the first bit is 0. $\mathcal{B}$ is the event that the second bit is 0.
- A card is drawn from a standard deck of cards. $\mathcal{A}$ is the event that the card's suit is hearts. $\mathcal{B}$ is the event that the cards rank is King.
- Two fair dice are rolled. $\mathcal{A}$ is the event that the first die lands an odd number. $\mathcal{B}$ is the event that the sum of the two dice is an odd number. This is where independence is not "clear in the English sense".

✍

Exercise: *Here are some examples of events – figure out which are dependent and which are independent. Check your intuition by really figuring out* $\mathbf{Pr}[\mathcal{A} \cap \mathcal{B}]$, $\mathbf{Pr}[\mathcal{A}]$, *and* $\mathbf{Pr}[\mathcal{B}]$.

- *A box contains three red balls and three blue balls. We first pick a ball at random and throw it away in the ocean. We then pick a second ball at random. $\mathcal{A}$ is the event that the first ball is blue, and $\mathcal{B}$ is the event that the second ball is blue.*
- *A box contains three red balls and three blue balls. We first pick a ball at random and throw it* back in the box. *We then pick a second ball at random. $\mathcal{A}$ is the event that the first ball is blue, and $\mathcal{B}$ is the event that the second ball is blue.*
- *We take a random permutation of the numbers $\{1, 2, 3, \ldots, n\}$. $\mathcal{A}$ is the event that the number 1 lands in the first place of this random permutation. $\mathcal{B}$ is the event that the number 2 lands in the second place of this random permutation.*