CS 30: Discrete Math in CS (Winter 2020): Lecture 14

Date: 3rd February, 2020 (Monday)
Topic: Probability: Conditional Probability, Independent
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. **Conditional Probability.** Often, we are interested in questions of the form

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

   Concrete examples:

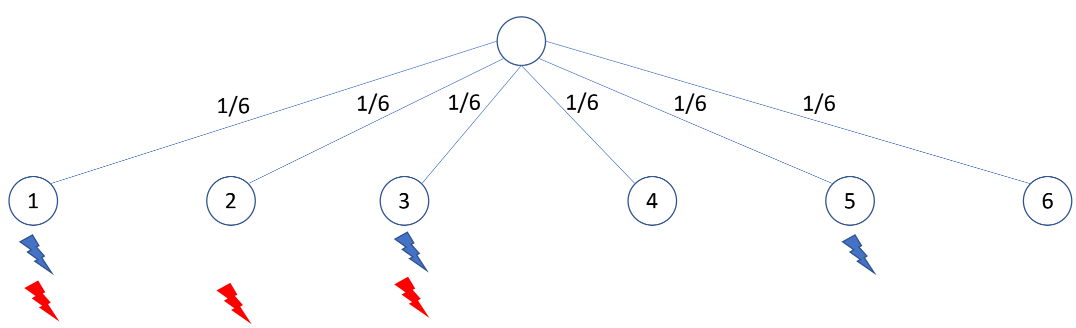
   - What is the probability that a roll of fair die lies in the set $\{1, 2, 3\}$ given that the roll is an odd number?
   - What is the probability that a roll of two fair dice sums to 6 given that the sum is an even number?

   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 $\mathcal{B}$. Furthermore, the whole set $\mathcal{A}$ is no longer the event we are interested in: indeed, all outcomes in the set $\mathcal{A} \setminus \mathcal{B}$ are irrelevant as they lie outside our new sample space. For example, in the example above, the event $\{2\}$ is irrelevant since 2 is not odd. Thus, the new event in this new sample space is $\mathcal{A} \cap \mathcal{B}$ — the part of $\mathcal{A}$ that lies in $\mathcal{B}$. Therefore, the new probability is calculated as:

$$\mathbf{Pr}[\mathcal{A} \mid \mathcal{B}] := \frac{\mathbf{Pr}[\mathcal{A} \cap \mathcal{B}]}{\mathbf{Pr}[\mathcal{B}]} \tag{Cond Prob}$$

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 die gives a number in $\{1, 2, 3\}$ when given that the die 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 die is odd. $\mathcal{E}$ is the event that the sum of the two dice is odd. What is $\mathbf{Pr}[\mathcal{A} \mid \mathcal{E}]$?*

2. **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

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

Here's an example 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?*

3. **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 $\mathcal{B}$,

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

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 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,

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

✍

Exercise: *Redo the above example using tree diagrams. To verify and also to get used to the fact that they are all the same.*

In fact, there are two successive generalizations of the law of total probability 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.*

4. **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 die rolls a 3. Let $\mathcal{B}$ be the event that the second die 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,*

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

✍

Exercise:*If $\mathcal{A}$ and $\mathcal{B}$ are independent, show that $\neg\mathcal{A}$ and $\neg\mathcal{B}$ are independent.*

Remark: *N events $\mathcal{E}_1, \mathcal{E}_2, \ldots, \mathcal{E}_N$ are mutually independent independent if*

$$\mathbf{Pr}[\mathcal{E}_1 \cap \cdots \cap \mathcal{E}_N] = \mathbf{Pr}[\mathcal{E}_1] \cdot \mathbf{Pr}[\mathcal{E}_2] \cdots \mathbf{Pr}[\mathcal{E}_N]$$

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.

The last three events may not be "clear" they are independent since they talk about the same event (same random string, same random card, same tuple of outcomes). Do the exercise to confirm that they are indeed independent.

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.*

5. **The Union Bound.** *We didn't do this on Monday, but may explore this later in the week. But it should be readable and understandable.*

Let us finish the lecture notes with a question related to the last lecture notes

> *Among 50 random people, what's the chance that at least one of them has a birthday on Jan 1?*

Once again, the assumption is that every person is equally likely to have a birthday on any of the 365 days.

To do so, let us set the stage by defining *smaller* events which will help us answer our question. Let us number the $N = 50$ people 1 to N. Let us say the event $\mathcal{E}_i$ occurs if the person i's birthday is Jan 1. Note two things:

$$\mathbf{Pr}[\mathcal{E}_i] = \frac{1}{365} \quad \text{for all } 1 \leq i \leq N \tag{1}$$

This is the assumption we are making. And the event $\mathcal{E}$ we are really interested in is

$$\mathbf{Pr}[\mathcal{E}_1 \cup \mathcal{E}_2 \cup \cdots \cup \mathcal{E}_N] \tag{2}$$

Now, if $\mathcal{E}_i$'s were mutually exclusive (as subsets of the sample space, these events were disjoint), then we would indeed have the probability of the union is the sum of the individual probabilities. However, that is not the case. $\mathcal{E}_1\mathcal{E}$ may not be the empty set since both people could be born on Jan 1. And thus, in general, $\mathbf{Pr}[\mathcal{E}_1 \cup \cdots \cup \mathcal{E}_N] \neq \sum_{i=1}^{N} \mathbf{Pr}[\mathcal{E}_i]$.

Nevertheless, the union's size can never be **larger** than the sum of the individual sizes, and thus we get an *upper bound* on the probability of the union. This upper bound is so widely used, that it has its own name. We may even see one use before we are done.

Theorem 5 (The Union Bound)**.** For any events $\mathcal{E}_1, \mathcal{E}_2, \cdots, \mathcal{E}_N$, we have

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

Therefore, the chance that one out of 50 people is born on Jan 1 is *at most* $50/365$. Also, another way we could have guessed the answer wouldn't exactly be the sum, is that there was nothing special about 50 people. However, if we had 400 people, the probability could not have been $400/365$ — probabilities **can never** exceed 1.

The above union bound is nice but it doesn't help us answer the question exactly. However, independence will allow us to solve this problem.

The main observation is that $\mathcal{E}_i$'s are mutually independent since they are random people with different independent sources of randomness (At some level, this is an assumption baked into the question.). Indeed, $\neg\mathcal{E}_1, \neg\mathcal{E}_2, \cdots, \neg\mathcal{E}_N$ are independent. Thus, $\mathbf{Pr}[\neg\mathcal{E}_1 \cap \cdots \cap \neg\mathcal{E}_N] = \prod_{i=1}^{N} \mathbf{Pr}[\neg\mathcal{E}_i] = \left(1 - \frac{1}{365}\right)^N$. Finally, we use De Morgans to get if $\mathcal{E} = \mathcal{E}_1 \cup \cdots \cup \mathcal{E}_N$, then

$$\neg\mathcal{E} = \neg\mathcal{E}_1 \cap \cdots \neg\mathcal{E}_N$$

in turn implying

$$\mathbf{Pr}[\mathcal{E}] = 1 - \mathbf{Pr}[\neg\mathcal{E}] = 1 - \left(1 - \frac{1}{365}\right)^N$$

which, if $N = 50$, is around 12.8%, that is, closer to $46.78/365$.