COMPSCI 230: Discrete Mathematics for Computer Science	April 10, 2019
Lecture 23	
Lecturer: Debmalya Panigrahi	*Scribe: Kevin Sun*

1 Overview

In this lecture, we will look at the fundamentals of probability theory. We begin with a classic problem whose result, at first glance, might seem counterintuitive. Afterwards, we formalize the our intuition of probability and present multiple applications.

2 The Monty Hall Problem

The *Monty Hall problem* is a classic problem that highlights the limitations of intuitive reasoning in probability theory. The problem goes as follows: you are on a game show, and there are three closed doors A, B, C. Behind one door is a car, and behind the other two are goats. Each door is equally likely to hide the car. You want the car, but you don't know which door to pick, so you go with A. The game host, who knows which door hides the car, then reveals a goat behind B. At this point, should you switch to C, or stick with A? Maybe the odds are now $50/50$, so does it even matter?

Before proceeding, we state the rules of the game more formally. In general, the host will always open a door that contains a goat, and the host never opens the door you initially pick. If the door you pick hides a goat, then the host only has one choice of door to open. If the door you pick hides the car, then the host randomly picks each of the other two doors with equal probability.

At first glance, one probably thinks that switching doesn't matter: after the host opens B, there are two remaining doors. We still can't decide between A or C with certainty, so we might as well stick with A. However, this intuition is incorrect: as we shall see, to maximize the probability of winning the car, you should *always* switch.

So our objective is to compare the following three strategies: "keep A" and "always switch". Recall that when the game is set up, each door is equally likely to hide the car. So let us evaluate the performance of each strategy over the long run, that is, if we played many games in a row using each strategy:

- Keep A: Since A hides the car $1/3$ of the time, this strategy wins roughly $1/3$ of the games.
- Always Switch: If A hides the car (which happens roughly $1/3$ of the time), then this strategy loses. However, if B or C hides the car, then the host opens C or B respectively. In these cases, switching wins! Thus, this strategy loses roughly $1/3$ games, so it wins roughly $2/3$ games.

As we can see, the "always switch" strategy performs better over the long run. In the subsequent sections, we will formalize our intuitive understanding of probability and apply it to a variety of problems.

3 Formalizing Probability

The foundations of mathematical probability lie in set theory, which at this point, we already understand fairly well. In general, when reasoning about probability, we must consider the set of all possible outcomes and the likelihood of each one; the formal definitions are given below.

Definition 1. *A* sample space *is a non-empty countable set. An* outcome *is an element of a sample space, and an* event *is a subset of the sample space (i.e., a set of outcomes). If A is an event of a sample space S, then we let $\overline{A} = S \setminus A$ denote the* complement *of A.*

Now recall that if a, b are real numbers and $a \leq b$, then $[a, b]$ denotes the set $\{x \in \mathbb{R} : a \leq x \leq b\}$.

Definition 2. *If S is a sample space, then a* probability function *on S is a total function $\Pr : S \to [0, 1]$ that satisfies $\sum_{x \in S} \Pr(x) = 1$. If $x \in S$ is an outcome, then $\Pr(x)$ denotes the* probability *of x. If $A \subseteq S$ is an event, then $\Pr(A)$ is defined as $\sum_{x \in A} \Pr(x)$ and denotes the probability of A.*

As we can see, a probability function specifies a value $\Pr(x) \in [0, 1]$ for every outcome $x \in S$. Furthermore, the requirement $\sum_{x \in S} \Pr(x) = 1$ is equivalent to $\Pr(S) = 1$; this formalizes our intuition that no matter what happens, the outcome will definitely be an element of S.

3.1 Probability Rules

Now that we have formalized the definitions related to probability, we can start studying some fundamental rules that allow us to reason about probabilistic events. Intuitively, we can reason about these rules as follows: the sample space S is a large dartboard, and an event A is a small region of the dartboard. Then the value of $\Pr(A)$ represents the ratio of the area of A to the area of S. Throughout the following, we let A and B denote events, i.e., subsets of a sample space S.

1. **Sum Rule:** If A and B are disjoint events, then $\Pr(A \cup B) = \Pr(A) + \Pr(B)$.
2. **Inclusion-Exclusion:** For any events A and B, $\Pr(A \cup B) = \Pr(A) + \Pr(B) - \Pr(A \cap B)$. In our dartboard analogy, this corresponds to finding the area of $A \cup B$, and so the rule follows from the principle of inclusion-exclusion that we saw in our lecture on combinatorics. Notice that the sum rule is a special case of inclusion-exclusion: if A and B are disjoint, then $\Pr(A \cap B) = 0$. We now discuss two other special cases of inclusion-exclusion.
   (a) **Boole's inequality:** $\Pr(A \cup B) \leq \Pr(A) + \Pr(B)$. This follows directly from inclusion-exclusion because $\Pr(A \cap B) \geq 0$.
   (b) **Union bound:** For any set of events, the probability of their union is at most the sum of the probability of each event. In other words, if $A_1, A_2, \ldots$ are events, then
   $$\Pr(A_1 \cup A_2 \cup \cdots) \leq \Pr(A_1) + \Pr(A_2) + \cdots .$$
   Note that this bound holds for a finite set of events, as well as an infinite set.
3. **Difference Rule:** $\Pr(A \setminus B) = \Pr(A) - \Pr(A \cap B)$. This rule follows by observing that any event A can be partitioned into the disjoint events $A \setminus B$ and $A \cap B$.
   (a) **Complement Rule:** $\Pr(\overline{B}) = 1 - \Pr(B)$. This rule follows from the difference rule by setting $A = S$, in which case $A \setminus B = \overline{B}$ and $\Pr(A) = \Pr(S) = 1$.

(b) **Monotonicity Rule:** If $A \subseteq B$, then $\Pr(A) \leq \Pr(B)$. This rule is known as monotonicity because it states that the Pr function does not decrease if we add outcomes to the event.

In everyday life, we often use these rules without noticing. However, as we have seen, informal justifications of probability can lead to incorrect results, so a formal mathematical understanding of these rules is critical.

3.2 Probabilistic Reasoning

Now that we have a list of rules, we can state a high-level strategy for reasoning about probabilistic events. Whenever faced with a probability problem, one should consider the following strategy:

1. Determine the sample space (i.e., the set of all possible outcomes).

2. Define the interesting event (i.e., the subset containing all interesting outcomes).

3. Calculate the probability of each outcome.

4. Calculate the probability of the event.

In this section, we will apply this strategy in several scenarios.

Rolling a Fair Die: Suppose we roll a fair 6-sided die once; let X denote the outcome of this roll. Here, the sample space is simply $\{E_1, E_2, E_3, E_4, E_5, E_6\}$, where E_i denotes the outcome "$X = i$". Since the die is fair, each outcome occurs with probability $1/6$; in other words,

$$\Pr(E_1) = \Pr(E_2) = \cdots = \Pr(E_6) = \frac{1}{6}.$$

Now let us formally calculate the probability that X is even. If we let A denote this event, then $A = \{E_2, E_4, E_6\}$. Since each outcome in A occurs with probability $1/6$ and all outcomes are disjoint, we can apply the sum rule to obtain $\Pr(A) = 3/6 = 1/2$, which matches our intuition.

Now consider the event containing outcomes where X is even or prime. Let B denote this event, and notice that there are multiple ways of calculating B:

1. Sum Rule: Since $B = \{E_2, E_3, \ldots, E_6\}$ and all outcomes are disjoint, we can conclude $\Pr(B) = \Pr(E_2) + \Pr(E_3) + \cdots + \Pr(E_6) = 5/6$.

2. Complement Rule: Notice that $\overline{B} = \{E_1\}$, so $\Pr(\overline{B}) = 1/6$. This implies $\Pr(B) = 1 - 1/6 = 5/6$.

3. Inclusion-exclusion: Recall that B must capture the event that X is even or prime. Therefore, $B = V \cup P$ where $V = \{E_2, E_4, E_6\}$ (evens) and $P = \{E_2, E_3, E_5\}$ (primes), and so $V \cap P = \{E_2\}$. By inclusion-exclusion, we have $\Pr(B) = \Pr(V) + \Pr(P) - \Pr(V \cap P) = 3/6 + 3/6 - 1/6 = 5/6$.

Rolling an Unfair Die: Now let's roll another 6-sided die, and let Y denote the outcome of this roll. The sample space is still $\{E_1, \ldots E_6\}$, where E_i denotes the outcome "$Y = i$". However, unlike the previous die, this die is not fair: for every $i \in \{1, 2, \ldots, 5\}$, this die is twice as likely to roll i than $i + 1$. In other words, the die obeys the following probability function:

$$\Pr(E_i) = 2 \cdot \Pr(E_{i+1}) \quad \forall i \in \{1, \ldots, 5\}.$$

Since exactly one of the 6 outcomes must still occur, and all of the outcomes are disjoint, the probability distribution still obeys the following equality:

$$\Pr(E_1) + \Pr(E_2) + \Pr(E_3) + \Pr(E_4) + \Pr(E_5) + \Pr(E_6) = 1.$$

Now let us first calculate the probability of each of the 6 outcomes. Notice that if we let $p = \Pr(E_6)$, then $\Pr(E_5) = 2p$, and similarly, $\Pr(E_4) = 2 \cdot \Pr(E_5) = 4p$. This line of reasoning yields

$$32p + 16p + 8p + 4p + 2p + p = 1,$$

and solving this equation yields $p = 1/63$. Now we can solve for the probability of each outcome. In particular, $\Pr(E_1) = 32/63$, which is much larger than $1/6$. Furthermore, the probability that Y is even is now $\Pr(E_2) + \Pr(E_4) + \Pr(E_6) = 16/63 + 4/63 + 1/63 = 21/63 = 1/3$. Similarly, we can see that the probability that Y is even or prime is (using the complement rule) $1 - \Pr(E_1) = 31/63$.

The Birthday Paradox: We now study a phenomenon known as the *birthday paradox*. This actually isn't a paradox in the strictest sense of the word, because the result we derive will be mathematically rigorous. However, for somebody who has never seen the result, it may sound quite surprising.

The setup is the following: assume that there are n students in a class, and a year has d days. Of course, we know that $d = 365$, but by using the variable d, our analysis can apply to a more general setting (e.g., if we set $d = 30$, then this is solving the problem of only considering students born in April). Assuming that each student is equally likely to have been born on any of the d days, what is the probability that all n birthdays are distinct?

Let's fix an ordering of the students, and count the number of outcomes. In this case, an outcome is a sequence of length n, and each element is one of the d days. Thus, the total number of possible outcomes is d^n. Furthermore, since the birthdays are all independent from each other and identical, every outcome is equally likely. Thus, each outcome occurs with probability $1/d^n$.

Now let D denote the outcome that the birthdays are distinct. For the birthdays to be distinct, the first student can have any one of d birthdays, but then the second birthday only has $(d-1)$ possibilities. This reasoning continues until the n-th student only has $(d-(n-1))$ possibilities. Thus, the total number of outcomes with no repeated birthdays is $d(d-1)(d-2)\cdots(d-(n-1))$.

Since each outcome is equally likely, the probability of our outcome having distinct birthdays is the following:

$$\begin{aligned}\Pr(D) = \frac{d(d-1)(d-2)\cdots(d-(n-1))}{d^n} &= \frac{d}{d}\cdot\frac{d-1}{d}\cdot\frac{d-2}{d}\cdots\frac{d-(n-1)}{d}\\ &= \left(1-\frac{0}{d}\right)\left(1-\frac{1}{d}\right)\left(1-\frac{2}{d}\right)\cdots\left(1-\frac{n-1}{d}\right).\end{aligned}$$

We now make use of the bound $1 - x < e^{-x}$ for any positive real number x. (This can be proved by, say, considering the Taylor series for e^{-x}.) Applying this inequality to each term above yields the following:

$$\begin{aligned}\Pr(D) &< e^{-0/d}\cdot e^{-1/d}\cdot e^{-2/d}\cdots e^{-(n-1)/d}\\ &= e^{-(\sum_{i=1}^{n-1} i/d)}\\ &= e^{-n(n-1)/(2d)}.\end{aligned}$$

Notice that as n increases, this upper bound on $\Pr(D)$ decreases. Intuitively, this makes sense: as the number of students increases, the probability that all birthdays are distinct decreases. In fact, by the pigeonhole principle, if $n \geq d+1$ then $\Pr(D) = 0$.

Equipped with this bound, we can determine the value of n that is large enough to ensure $\Pr(D) < 1/2$. It is straightforward to verify that if $n \geq 25$, then $e^{-n(n-1)/(2\cdot 365)} < 0.44$. This means that in a class of only 25 students, it is more likely than not that two students share a birthday!

3.3 Conditional Probability

Finally, we formalize the notion of conditional probability.

Definition 3. *Let A and B be two events of a sample space. Then the probability of the event "A given B", denoted $A|B$, is defined as the following:*

$$\Pr(A|B) = \frac{\Pr(A \cap B)}{\Pr(B)}.$$

Intuitively, we can appreciate the utility of conditional probabilities by considering the following scenario: Suppose B is an extremely unlikely event, i.e., $\Pr(B)$ is very small, and A is a strict subset of B, so $\Pr(A)$ is even smaller. Normally, we do not expect A to happen; however, if we were told that B happened, we start to suspect A also happened. This is given by the probability of A conditioned on the fact that B happened.

We can also extend the "dartboard" analogy to this situation. In this case, we know that the dart hit the region corresponding to B, and we want to know if the dart also hit A. Then we only need to concern ourselves with the region of A that is contained in B, and the new "sample space" is only the region corresponding to B.

Independence: Intuitively, two events A and B are independent if knowing the outcome of B does not affect the probability of A. Mathematically, A is *independent* of B if $\Pr(A|B) = \Pr(A)$. Notice that, from the definition of conditional probability, this equation is equivalent to $\Pr(A \cap B) = \Pr(A) \cdot \Pr(B)$. From this perspective, it is clear that A is independent of B if and only if B is independent of A.

4 Summary

In this lecture, we introduce the basics of probability theory giving the relevant definitions and rules. We also applied a four-step reasoning process to multiple examples, including the birthday paradox, and concluded with a brief look at conditional probability.