Lecture 2: Conditional Probability

Rajat Mittal

IIT Kanpur

Armed with the basic definitions of probability theory, let us dive into the concept of *conditional probability*. The main motivation behind conditional probability is, occurence of an event A *can potentially* influence the probability of another event B. For instance, a random person in Kanpur might be ill with a certain probability. Though, if we know that the person is presently at a hospital, we would guess that the probability of being ill should substantially increase. In this case, A will be the event that person is ill and B being the event that the person is in the hospital.

To capture this kind of scenario, we introduce *conditional probability* in this lecture note. Later part of this note will discuss Bayes theorem, one of the most fundamental results in conditional probability. We will end with some well known examples/puzzles in the world of conditional probability.

1 Conditional probability

How should we define probability of an event A given that an event B has happened?

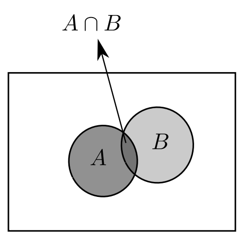

Fig. 1. Two events A and B

Exercise 1. Look at Fig. 1, what should be the probability of event A given that event B has happened?

Given two events A, B, the conditional probability of A given B is defined by,

$$P\left(\frac{A}{B}\right) := \frac{P(A \cap B)}{P(B)}.$$

You can read $P\left(\frac{A}{B}\right)$ as the probability of event A *given* event B.

Note 1. This is how we have *defined* conditional probability and not derived it. Though, the definition matches our intuition.

To take a simple example, what is the probability of getting two heads given that at least one coin toss was heads. The sample space consists of 4 elements. Define event A to be the set of outcomes with two heads. Define B to be the set of outcomes with at least one heads.

$$P\left(\frac{A}{B}\right) = \frac{P(A \cap B)}{P(B)} = \frac{1/4}{3/4} = 1/3.$$

Sometimes, it is easy to compute the proabability of an event by dividing the sample space in to disjoint parts. A kid has two machines at her disposal, red and blue. Since blue is her favorite color, she picks that machine twice as much as the red one. We also know that half the balls from red machine are defective and one-quarter balls from blue machine are defective. If the kid obtains a new ball, what is the probability that the ball is defective?

There are two cases.

– Picked ball is from red machine (1/3 probability). Then, 1/2 probability to get a defective ball given it is from red machine. So, by definition of conditional probability, we get a defective ball with probability $1/3 \times 1/2 = 1/6$.
– Picked ball is from blue machine (2/3 probability). Then, 1/4 probability to get a defective ball given it is from blue machine. So, by definition of conditional probability, we get a defective ball with probability $2/3 \times 1/4 = 1/6$.

Exercise 2. Let A be the event that the ball is defective and B_r, B_b be events that the ball is from red/blue machine respectively. For all events whose probability is calculated above, write these events in terms of A, B_r and B, b.

The total probability is the sum of two terms, getting a red machine ball being defective and getting a blue machine ball being defective, and is equal to 1/3.

This law can be generalized (call it *partition formula*) and you will prove it in the assignment. Given event A and *disjoint* partition $B_1, B_2, \cdots, B_m$ of sample space,

$$P(A) = \sum_{i=1}^{m} P(B_i) P\left(\frac{A}{B_i}\right).$$

Consider the following situation: given n sticks of different length and n holes in a line, what is the probability that the stick at the k-th hole is visible from the left? The main issue here is to find events B_i, clearly A is the event that the stick at k-th hole is visible.

Our first attempt could be, B_i is the event that i-th stick (when arranged in order of height) goes in the k-th hole. In this case, $P(B_i)$ is $1/n$. What is $P\left(\frac{A}{B_i}\right)$? You will show in the assignment,

$$P\left(\frac{A}{B_i}\right) = \frac{1}{(n-1)!}\binom{i-1}{k-1}(k-1)!(n-k)!.$$

Exercise 3. Calculate the probability of event A.

Though, a different partition will give the solution directly. Given a subset $S \subseteq [n]$ of size k, let B_S be the event that the sticks from set S go to first k holes. Now $P\left(\frac{A}{B_S}\right)$ is simply $1/k$, applying the partition formula,

$$P(A) = \sum_S P(B_S) P\left(\frac{A}{B_S}\right) = 1/k \sum_S P(B_S) = 1/k.$$

How did we get the last equality?

1.1 Independence

Exercise 4. What is the probability that the two throws of a dice give the same number? What if we know that the number displayed on the first dice is prime? What if we know that the sum of numbers displayed is prime? What if we know that the sum is less than 5?

We notice the two extremes; if sum of numbers is prime then they can't be equal. On the other hand, the number on the first dice being prime has no effect on them being equal. The later situation can be thought as, event A of the first number being prime is completely independent to event B of them being equal. Again, we formalize this notion with the concept of *independence* between events.

Let us take few more examples.

1. What is the probability of obtaining a head again if we knew that the first toss gave a head, while tossing an unbiased coin twice.
2. Suppose Euler misses school on a day with probability $1/2$. What is the probability that he misses school twice on two consecutive days?

Given that the coin is unbiased in the first example, the outcome of second toss is independent of the first toss. On the other hand, if Euler misses the school on first day, he might also miss it the next day with high probability (because he might be out of station or ill). In this sense, the event that Euler is absent on the first day is not independent of the event that he misses the school on the second day.

Our intuition suggests that event A and B should be independent if $P(A)$ and $P\left(\frac{A}{B}\right)$ are equal. We define two events A, B to be independent if,

$$P(A \cap B) = P(A)P(B) \Leftrightarrow P(A) = P\left(\frac{A}{B}\right).$$

Suppose there are two red and two blue balls in a bin. You pick two balls out of it sequentially, one after another. Let R_1 be the event that the first draw is a red ball. Simiarly, let B_2 be the event that second ball is blue.

Exercise 5. Are these two events, R_1 and B_2, indpendent?

There can be two ways to perform this experiment. In the first case, you can pick a ball, replace it, and then draw the second ball.

Exercise 6. Show that R_1 and B_2 are independent in this case.

In the other case, you do not replace the ball. It seems that the probability of event B_2 should increase if the first ball is red (event R_1 occured). Intuitively, they should be dependent. Let us check it.

The conditional probability $P\left(\frac{B_2}{R_1}\right)$ is $2/3$ (since two blue ball and one red ball remains). What is the probability of second ball being blue?

Exercise 7. Show that $P(B_2)$ is $1/2$, significantly lower than the conditional probability. You can calculate it using partition formula or by using symmetry.

The difference between *disjoint* events and *independent* events can be confusing. Remember, if A and B are independent, occurence of A has no effect on occurence of B. On the other hand, if A and B are disjoint, occurence of A rules out the occurence of B (they are *highly* dependent).

2 Bayes' theorem

Suppose a scientific theory predicts that there will be a solar eclipse on 1st Oct. with high probability (say p). If you observe that there is a solar eclipse on 1st Oct, what is the probability that the theory is correct? Do we have enough information to answer this question? Such problems are called hypothesis testing.

We can frame this problem in terms of events. Define event A to be *scientific theory being true* and event B to be that *solar eclipse happens on 1st Oct.* The conditional probability $P\left(\frac{B}{A}\right)$ is given to be p, and we want to calculate the conditional probability $P\left(\frac{A}{B}\right)$.

Exercise 8. What should this probability, $P\left(\frac{A}{B}\right)$, depend upon? Is the value of $P\left(\frac{B}{A}\right)$ sufficient to find the value $P\left(\frac{A}{B}\right)$?

Your first guess might be, if $P\left(\frac{B}{A}\right)$ is high then $P\left(\frac{A}{B}\right)$ should be high too. Consider the following scenario, it is certain that solar eclipse is going to happen on 21st Oct., irrespective of the prediction of the theory. Though, the probability of event A, scientific theory being true, is pretty small. In this case, $P\left(\frac{B}{A}\right)$ will be very high, but it should not have any relevance about the truthfulness of the scientific theory. Summarizing, the base probabilities of the events, $P(A)$ and $P(B)$, are also important in the calculation of $P\left(\frac{A}{B}\right)$.

We fall back to the formula for conditional probability. The probability $P\left(\frac{A}{B}\right)$ is the ratio of $P(A \cap B)$ and $P(B)$. Expressing $P(A \cap B)$ as a product of $P\left(\frac{B}{A}\right)$ and $P(A)$ gives us *Bayes theorem.*

Theorem 1. *Bayes: Let A and B be two events. Then the conditional probability $P\left(\frac{A}{B}\right)$ is given by,*

$$P\left(\frac{A}{B}\right) = \frac{P\left(\frac{B}{A}\right) Pr(A)}{Pr(B)}.$$

Make sure that you can prove this theorem using the definition of conditional probability. The denominator in Bayes theorem is mostly obtained using the formula,

$$P(B) = P\left(\frac{B}{A}\right) Pr(A) + P\left(\frac{B}{A^c}\right) Pr(A^c).$$

Below, you will see the application of Bayes' formula in various settings.

- In Mumbai, 90% of the taxis are black and the rest are white. It was observed by Times of India that white taxi drivers are very rash and are 5 times more likely to be involved in an accident as compared to a black taxi. Recently there was an accident involving a taxi, what is the probability that the taxi was white?
  Define A to be the event that there was an accident. Define C_W and C_B to be the events that the taxi is white (respectively black). The quantity of interest is $P\left(\frac{C_W}{A}\right)$. Applying Bayes' formula,

  $$P\left(\frac{C_W}{A}\right) = \frac{P\left(\frac{A}{C_W}\right) P(C_W)}{P\left(\frac{A}{C_W}\right) P(C_W) + P\left(\frac{A}{C_B}\right) P(C_B)}.$$

  We don't know these probabilities explicitly, but the ratio $\frac{P(C_W)}{P(C_B)}$ and $\frac{P\left(\frac{A}{C_W}\right)}{P\left(\frac{A}{C_B}\right)}$ are known.

  Exercise 9. Calculate the probability that the taxi was white, given that it was involved in an accident.

- One of the main application of Bayes theorem is in medical diagnosis. Assume that there is a test for early detection of cancer and a study shows that it is very successful. Specifically, the study shows that if a person has cancer then the test will diagnose cancer with probability .9 (even though only 1% people have cancer in general population). Similarly, if a person does not have cancer then the test correctly diagnoses with probability .9.
  Suppose a person is tested and the test shows that the person has cancer, what is the probability that the person actually has cancer? Make an educated guess, without calculating the probability using Bayes theorem. A naive guess would be that the test works in both cases, so it seems pretty accurate. Hence, the person has cancer with very high probability.
  Now, we apply Bayes theorem to calculate the probability. Say A be the event that person has cancer and B be the event that test outputs that person has cancer. The probability of event A is pretty low, $P(A) = .01$ (this is known as *base rate*). The probability of event B can be calculated by considering two disjoint events A and A^c. Applying Bayes' formula,

  $$P\left(\frac{A}{B}\right) = \frac{.9Pr(A)}{.9Pr(A) + .1Pr(A^c)} \approx .1.$$

This seems like a big surprise. Even though the probability of the person having cancer has increased from 0.01 to .1, still it is much smaller than the success probability of the test (.9). It shows that base rate matters a lot in this calculation and should not be ignored.

- Another area of application is machine learning, we take a toy example. A spam detector, using data from a user, has found these patterns.
  1. Word *money* appears in 50% of spam emails.
  2. There is a particular friend of the user who likes money, word *money* appears in 40% of his emails.
  3. Except him, word *money* appears in only 5% of non-spam emails.

  The user has received another email with word *money*, what is the probability that the email is a spam? Again, we need base rates to make this calculation. Assume that the special friend is responsible for 10% of the emails to the user. Rest of the emails can equally be spam or non-spam.
  Let M be the event that email contains word *money*. Let S, N, F be events that the email is a spam, non-spam and sent by friend respectively. Using Bayes theorem,

$$P\left(\frac{S}{M}\right) = \frac{P\left(\frac{M}{S}\right)P(S)}{P\left(\frac{M}{S}\right)P(S) + P\left(\frac{M}{N}\right)P(N) + P\left(\frac{M}{F}\right)P(F)}.$$

  Exercise 10. Calculate the probability $P\left(\frac{S}{M}\right)$.

2.1 Monty Hall problem

This famous problem is posed in the context of an American game show. It is more than 40 years old and still debated among many probability experts. Let us look at the problem directly, look at Wikipedia to read about the history and fights (about the problem) from Wikipedia.

A user is participating in a game show hosted by famous *Monty Hall*. There are 3 doors, one of them hides a car and other two have goats behind them. You are asked to pick a door, then the game show host opens one of the other two doors and reveals a goat. Remember that at least one of other two doors have a goat and host knows it. Assuming that you are not interested in a goat, should you switch the door?

Let me ask a simpler question first.

Exercise 11. We look at tweets of a politician. Suppose, the politician always tweets random useless stuff on Sunday. On the other hand, his Saturday tweets are useful half the time (with probability half they are useless). We assume that he tweets equally on Saturday and Sunday. You woke up one day and saw a useless tweet, is it more probable to be a Saturday or a Sunday? Try to answer this question on an intuitive level, without applying Bayes theorem.

Let us get back to our original question, Monty Hall problem. What will be your guess? Most of the people don't think that switching will affect their chances. This problem is very famous because of the counter-intuitive nature of the result.

Assume the standard assumptions that car could be behind any door with equal probability. Also if you pick the door with car, Monty will choose the door to be opened uniformly at random (out of other two). Suppose the doors are numbered $1, 2$ and 3. Without loss of generality, we can assume that you pick the door 1. Then, say Monty opens door 2.

We are interested in the conditional probability that the car is behind door 1, given that Monty opened door 2. Let us calculate the probability explicitly.

Let D_i be the event that car is behind door i, $P(D_i) = 1/3$. Let B be the event that Monty opens door 2. Then, we are interested in $P\left(\frac{D_1}{B}\right)$.

$$P\left(\frac{D_1}{B}\right) = \frac{P\left(\frac{B}{D_1}\right)Pr(D_1)}{P\left(\frac{B}{D_1}\right)P(D_1) + P\left(\frac{B}{D_2}\right)P(D_2) + P\left(\frac{B}{D_3}\right)Pr(D_3)}.$$

Note 2. Ideally, all probabilities should be with the event that you have picked door 1. Since it is common, we have chosen to skip it for brevity.

Exercise 12. Convince yourself that the formula is correct.

We know that $P\left(\frac{B}{D_2}\right)$ is 0 and $Pr(D_1) = Pr(D_2) = Pr(D_3) = 1/3$. So,

$$P\left(\frac{D_1}{B}\right) = \frac{P\left(\frac{B}{D_1}\right)}{P\left(\frac{B}{D_1}\right) + P\left(\frac{B}{D_3}\right)}.$$

The probability $P\left(\frac{B}{D_1}\right)$ is 1/2 because Monty could have chosen door 2 or door 3. Though $P\left(\frac{B}{D_3}\right)$ is 1 because Monty's only choice was to open the door 2. This tell us that $P\left(\frac{D_1}{B}\right) = 1/3$ and hence $P\left(\frac{D_3}{B}\right) = 2/3$. So, it was beneficial to switch the doors for you.

Exercise 13. Can you see the resemblance with tweet question asked earlier?

We can think of switching vs non-swtiching this way. If car was behind door 1, Monty would have opened door 2 with half probability. If car was behind door 3, Monty would have opened door 2 with probability 1. Given that he has opened door 2, it is more probable that the car was behind door 3.

3 Assignment

Exercise 14. For the events $A_1, A_2, \cdots, A_n$, prove,

$$P(\bigcup_{i=1}^{n} A_i) = \sum_{i=1}^{n} P(A_i) - \sum_{1 \le i < j \le n} P(A_i \cap A_j) + \cdots + (-1)^{n+1} P(\bigcap_{i=1}^{n} A_i).$$

Exercise 15. How many numbers are there between 1 and 1000 divisible by 2,3 or 5?

Exercise 16. Suppose 3% of the students in JNU are enrolled in the Math department. By looking at the personality sketch of Ramanujan, you think he is 4 times more probable to be in Math than to be in other departments. What is the probability that Ramanujan is in Math department.

Exercise 17. Read more about Monty Hall problem and its variations from Wikipedia.

Exercise 18. An event A is positively correlated to B if $P\left(\frac{A}{B}\right) \ge P(A)$. Suppose A is positively correlated to B, then show that,

- B is positively correlated to A.
- B^c is negatively correlated to A. What will be the definition of negatively correlated?

Exercise 19. Given event A and *disjoint* partition $B_1, B_2, \cdots, B_m$ of sample space, show that

$$P(A) = \sum_{i=1}^{m} P(B_i) P\left(\frac{A}{B_i}\right).$$

Exercise 20. Remember the stick problem in the text. A is the event that stick at the k-th hole is visible. B_i is the event that i-th stick goes at the k-th hole. Show,

$$P\left(\frac{A}{B_i}\right) = \frac{1}{(n-1)!}\binom{i-1}{k-1}(k-1)!(n-k)!.$$

Simplify the expression of $P(A)$ and show that it is $1/k$.

Exercise 21. Let there be a student Ramanujan from SRCC, Delhi. He has a very small circle of friends. According to them, he is an introvert and is known as "nerd". Some people speculate that he feels very lonely. It is known that only 3% of students in SRCC are from Math department. Sort the following options in increasing order of probability.

- Student of Math dept.
- Student of Commerce dept.
- He feels very lonely.
- He is from Commerce dept and has a minor in Math.

Frame this problem in terms of hypothesis testing.

References

1. D. Stirzaker. Elementary Probability. *Cambridge University Press*, 2003.
2. D. Kahneman. Thinking, Fast and Slow. *Farrar, Straus and Giroux*, 2011.