# 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, occurrence 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/applications/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? Assume $P(B) > 0$.

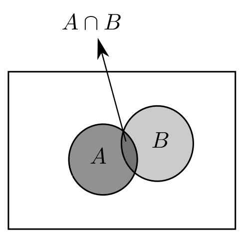


**Fig. 1.** Two events $A$ and $B$

*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,

$$P[A|B] := \frac{P(A \cap B)}{P(B)}.$$

You can read $P[A|B]$ 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. If you don't see it directly, look at the case when each element in sample space (assume the sample space is finite) has the same probability of occurrence.

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.

$$P[A|B] = \frac{P(A \cap B)}{P(B)} = \frac{1/4}{3/4} = 1/3.$$

Sometimes, it is easy to compute the probability 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 often 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 $1/3 \times 1/2 = 1/6$ (the probability that the kid picked the red machine AND the ball is defective).
- 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 $2/3 \times 1/4 = 1/6$ (the probability that the kid picked the blue machine AND the ball is defective).

*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,

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

*Exercise 3.* Is it necessary that $B_1, B_2, \cdots, B_m$ forms a disjoint partition. Show that it is enough if $A \subseteq \cup_i B_i$, where $B_i$ are disjoint.

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 $P\left[A|B_i\right]$? You will show in the assignment,

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

*Exercise 4.* 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 $P\left[A|B_S\right]$ is simply $1/k$, applying the partition formula,

$$P(A) = \sum_S P(B_S) P\left[A|B_S\right] = 1/k \sum_S P(B_S) = 1/k.$$

How did we get the last equality?

Let us take another example (from Stirzaker's book [2]). You enter a casino with $k$ Rs. At each time step you bet 1 Rs. You win with probability $p$ ($p < 1/2$ for a sensible casino) and get 2 Rs., otherwise you get 0. You leave if you earn $K$ Rs. ($K > k$) or your money finishes. What is the probability that you leave with nothing?

Denote $p_k$ to be the probability that you leave with nothing when you start with $k$ Rs. Let $W$ denote the event that you win on first try. This shows

$P(\text{you leave with nothing}) = P(W)P(\text{you leave with nothing}|W) + P(W^c)P(\text{you leave with nothing}|W^c)$.

With our notation,

$$p_k = p_{k+1}p + (1-p)p_{k-1}.$$

Simplifying a bit,

$$p_{k+1} - p_k = \frac{1-p}{p}(p_k - p_{k-1}).$$

This leads to

$$p_{k+1} - p_k = \left(\frac{1-p}{p}\right)^k (p_1 - 1),$$

since $p_0 = 1$. Again, using repeated addition,

$$p_k = 1 + (p_1 - 1)\frac{\left(\frac{1-p}{p}\right)^k - 1}{\left(\frac{1-p}{p}\right) - 1}.$$

*Exercise 5.* How can we get $p_1$?

That will be obtained by noticing that $p_K = 0$. We substitute $p_1$ in the previous equation and get

$$p_k = \frac{\left(\frac{1-p}{p}\right)^K - \left(\frac{1-p}{p}\right)^k}{\left(\frac{1-p}{p}\right)^K - 1}.$$

*Sampling:*

Suppose you want to estimate the number of prime numbers between 1 and $1,00,000$. One expensive way is to go through all the numbers and check how many of them are prime. If we can be happy with an estimate (approximate fraction), a smarter way is to pick a small collection of random numbers and check how many of them are prime. This method of getting information about a set (in this case the probability of being prime) by picking a smaller set is called *sampling*. It is one of the most basic and important method in the toolbox of statistics.

Let there be $t$ primes between 1 and $n = 1,00,000$, and we want to sample $k$ numbers out of it. There are two ways to sample in our problem.

- Sampling without replacement: In this case we pick a number, check if it is prime or not, and then throw it out. The probability that we get a prime number in the $i$-th pick depends on our earlier pick. For example, if the first pick was prime, then second will be prime with probability $(t-1)/(n-1)$. On the other hand, if the first pick was a non-prime then the second pick will be prime with probability $t/n-1$.
- Sampling with replacement: In this case, we pick numbers one by one and put it back into the original set before our next pick. In other words, in any pick the probability that we get a prime is always $t/n$, independent of what we picked earlier.

What is the probability that your second pick is prime. Using conditional probability,

$$P[\text{2nd prime}] = P[\text{2nd prime}|\text{1st prime}] \cdot P[\text{1st prime}] + P[\text{2nd prime}|\text{1st non-prime}] \cdot P[\text{1st non-prime}].$$

*Exercise 6.* Calculate this probability for both methods of sampling.

You see that conditional probability automatically arises in such cases. Let us talk about our original question. How do you get an estimate of $t$? If we get $l$ primes, then our estimate would be $l(n/k)$ for $t$. It can be shown that the probability of getting a good estimate of $t$ keeps increasing with larger $n, t, k$ (for both methods of sampling); we will do this later in the course. Depending on the situation, we sometimes use sampling with replacement and sometimes sampling without replacement. In many cases it can be shown that their output is very similar.

*Exercise 7.* What is the probability that the first $k/2$ picks are prime and then the remaining ones are non-prime (assume $t \geq k/2$). Calculate this for both methods of sampling. How did you obtain the answer, what concept of probability did you use?

How do we formally derive this probability? Say $A_i$ is the desired event for the $i$-th pick (it means that the picked number is prime for $i \leq k/2$ and non-prime for $i > k/2$). We need the probability for $P[A_1 \cap A_2 \cdots \cap A_k]$. From conditional probability,

$$P[A_1 \cap A_2 \cdots \cap A_k] = P[A_k | A_{k-1} \cap A_{k-2} \cdots \cap A_1] P[A_{k-1} \cap A_{k-2} \cdots \cap A_1].$$

This can be easily generalized to,

$$P[A_1 \cap A_2 \cdots \cap A_k] = P[A_k | A_{k-1} \cap A_{k-2} \cdots \cap A_1] P[A_{k-1} | A_{k-2} \cdots \cap A_1] \cdots P[A_1].$$

Each of the terms on the right hand side are easy to calculate. This is called *chain rule.*

*Exercise 8.* Find the answer for both methods of sampling.

You might see that sampling with replacement is easier to analyze as compared to sampling without replacement. The reason is that the picks are *independent* of each other, one pick doesn't interfere in another pick's output. When talking about conditional probability, the concept of *independence* arises quite naturally. One of the special case for conditional probability is $P(A|B) = P(A)$; intuitively, the event $B$ has no influence on event $A$. Using the definition of conditional probability, this is equivalent to,

$$P(A \cap B) = P(A)P(B).$$

Notice that the above definition is symmetric, i.e., if $B$ has no influence on $A$ then $A$ has no influence on $B$. In this case, event $A, B$ are called to be *independent.* This is a very important sub-case and arises quite naturally and frequently. We will discuss this in detail in future.

## 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 $P[B|A]$ is given to be $p$, and we want to calculate the conditional probability $P[A|B]$.

*Exercise 9.* What should this probability, $P[A|B]$, depend upon? Is the value of $P[B|A]$ sufficient to find the value $P[A|B]$?

Your first guess might be, if $P[B|A]$ is high then $P[A|B]$ 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, $P[B|A]$ 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 $P[A|B]$.

We fall back to the formula for conditional probability. The probability $P[A|B]$ is the ratio of $P(A \cap B)$ and $P(B)$. Expressing $P(A \cap B)$ as a product of $P[B|A]$ 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,*

$$P[A|B] = \frac{P[B|A]\,Pr(A)}{Pr(B)}.$$

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,

$$P(B) = P[B|A]\,Pr(A) + P[B|A^c]\,Pr(A^c).$$

Below, you will see some applications of Bayes' formula.

- 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 $P[C_W|A]$. Applying Bayes' formula,

  $$P[C_W|A] = \frac{P[A|C_W]\,P(C_W)}{P[A|C_W]\,P(C_W) + P[A|C_B]\,P(C_B)}.$$

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

  *Exercise 10.* 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,

  $$P[A|B] = \frac{.9Pr(A)}{.9Pr(A) + .1Pr(A^c)} \approx .1.$$

  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 word *money* appears in 40% of emails from bank.
  3. Excluding bank emails, 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 bank 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 $S, N, B$ be events that the email is a spam, non-spam and sent by bank respectively. Using Bayes theorem,

  $$P[S|M] = \frac{P[M|S]\,P(S)}{P[M|S]\,P(S) + P[M|N]\,P(N) + P[M|B]\,P(B)}.$$

  *Exercise 11.* Calculate the probability $P[S|M]$.

*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?

You can take the following standard assumptions.

- Car could be behind any door with equal probability.
- If you pick the door with car, Monty will choose the door to be opened uniformly at random (out of other two).
- 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.

Should you switch? Before worrying about this complicated looking problem, let us solve a simpler question first.

*Exercise 12.* 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 are the chances that Monty picked door 2 if the car was behind door 1? What are the chances that Monty picked door 2 if the car was behind door 3? What does it have to do with the simpler question asked? Don't worry if the connection is not clear. Next, we show a formal proof that switching helps.

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 $P[D_1|B]$.

$$P[D_1|B] = \frac{P[B|D_1]\,Pr(D_1)}{P[B|D_1]\,P(D_1) + P[B|D_2]\,P(D_2) + P[B|D_3]\,Pr(D_3)}.$$

*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 13.* Convince yourself that the formula is correct.

We know that $P[B|D_2]$ is 0 and $Pr(D_1) = Pr(D_2) = Pr(D_3) = 1/3$. So,

$$P[D_1|B] = \frac{P[B|D_1]}{P[B|D_1] + P[B|D_3]}.$$

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

*Exercise 14.* Can you see the resemblance with tweet question asked earlier now?

We can think of switching vs non-switching 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 15.* Suppose 3% of the students in JNU are enrolled in the Math department. The personality sketch of Ramanujan is 4 times more probable given that he is in Math as compared to when he is in other departments. What is the probability that Ramanujan is in Math department given his personality sketch.

*Exercise 16.* Read more about Monty Hall problem and its variations from Wikipedia.

*Exercise 17.* An event $A$ is positively correlated to $B$ if $P\left(\frac{A}{B}\right) \geq P(A)$. 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 18.* Given event $A$ and *disjoint* partition $B_1, B_2, \cdots, B_m$ of sample space, show that

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

*Exercise 19.* 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,

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

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

*Exercise 20.* 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.

The last problem and many such examples (where our intuition and probability are in conflict) are given in Kahneman's book [1].

## References

1. D. Kahneman. *Thinking, fast and slow.* Farrar, Straus and Giroux, 2011.
2. D. Stirzaker. *Elementary probability.* Cambridge University Press, 2003.