Lecture 3: Conditional Probability

Yunwen Lei

School of Computer Science, University of Birmingham

In this class, we will introduce the concept of conditional probability. This is a probability under a condition some other events have already happened.

1 Motivation

We first give an example to motivate the introduction of conditional probability.

- The chance of a random person has the flue is about 0.01

$$\text{Probability of flu}: \quad \mathbb{P}(\text{flu}) \approx 0.01.$$

  This is a prior probability. Normally, the probability of a person having a flue is low.

- Now suppose we are given some new information: you have a slight fever. Under this new information, it is reasonable to imagine that the chances of flu "increase"

$$\text{Probability of flu given fever}: \quad \mathbb{P}(\text{flu}|\text{fever}) \approx 0.4.$$

  New information changes the prior probability to the posterior probability, which is the updated probability after some observation. We can translate posterior as "After you getting the new information"

  $\mathbb{P}(A|B)$ is the (updated) conditional probability of A, given the information B

- Suppose we have more information: roommate has flu. Under this new information, it is almost sure that the person would have flue

$$\text{Probability of flu given fever, roommate flu}: \mathbb{P}(\text{flu}|\text{fever AND roommate flu}) \approx 1$$

In this example, we know that the new information changes our belief on some event to happen. Without information, the event is unlikely to happen. But given some other information, this event becomes almost certain. We need a mathematical way to explain this process. This is the conditional probability.

Roughly speaking, the conditional probability $\mathbb{P}(A|B)$ is related to this question: how does new information B affect the probability of an event A?

$$\mathbb{P}(A|B) = \text{The probability that } A \text{ occurs, given that } B \text{ HAS occurred}$$

Let us consider an example.

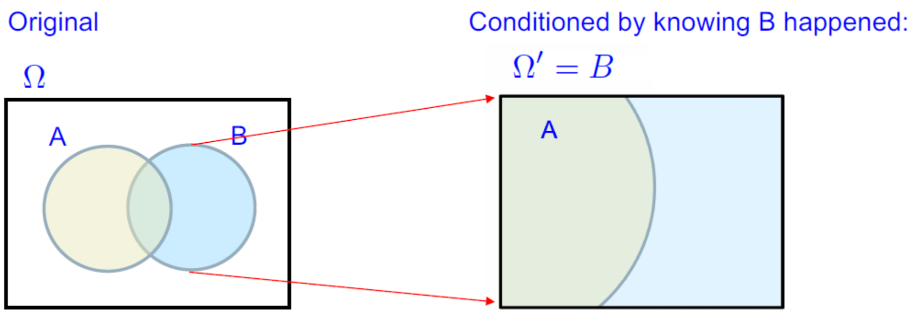
Example 1 (Toss of Two Dice)**.** Suppose I toss two dice, one in front of you, and the other where you cannot see the result. The one you can see shows 3 dots. What is the probability that more than 8 dots shown on two dice? Or: $\mathbb{P}(A|B)$, where

- A ="The total number of dots is more than 8"

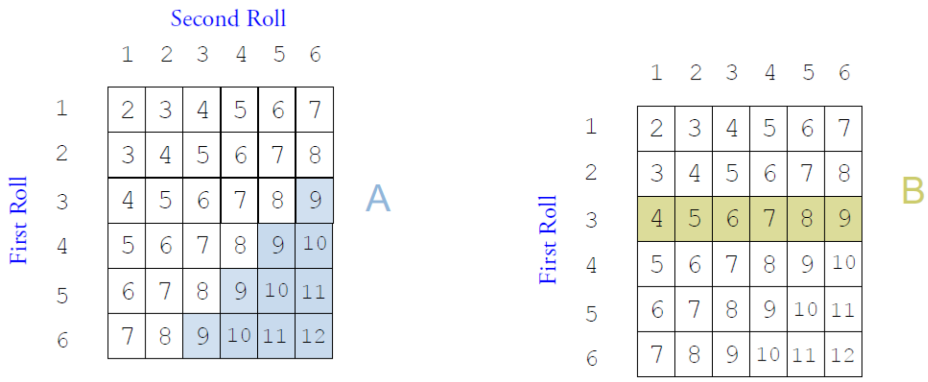
- $B =$"The first die shows 3 dots"

The key to solving such problems is to realize that there are two probability spaces:

- the one before you know whether B has happened or not, and
- the one that has been "conditioned" by knowing that B has definitely happened, so the sample space has shrunk and the proportion representing event A may change. In other words, we need to focus on the event B and define a probability law such that $\mathbb{P}(B|B) = 1$.



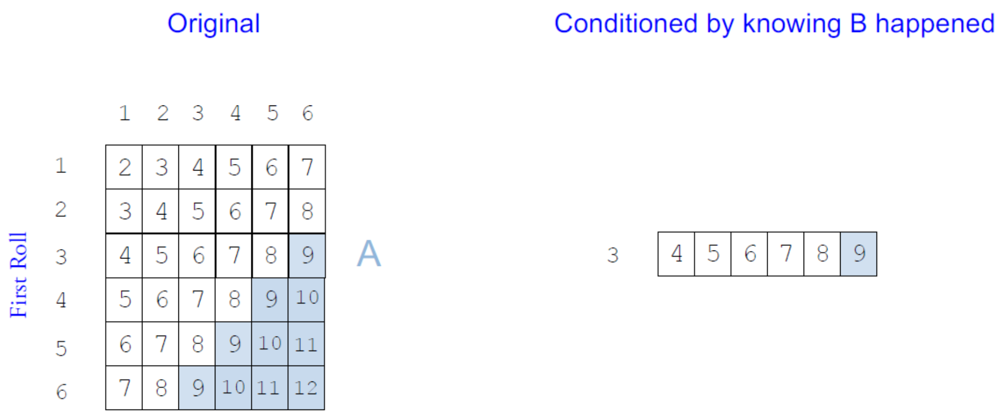
We use blue color to show the event A and yellow color to show the event B. Note that the event A has 10 outcomes, while the event B has 6 outcomes. The intersection has only one outcome $(3, 6)$, i.e., the first die shows 3 while the second die shows 6.



It is clear that

$$\mathbb{P}(A) = 10/36 \quad \mathbb{P}(B) = 6/36 \quad \mathbb{P}(A \cap B) = 1/36$$

Now we show what the conditional probability should intuitively be. Since we know the event B happens, we should update the sample space Ω to a new sample space $\Omega' = B$, which only has 6 outcomes. Since we are interested in the probability of A to happen, we need to consider the intersection $A \cap B = \{(3, 6)\}$. As the experiment has equally likely outcomes, it is reasonable to assign $1/6$ to the conditional probability, i.e., $\mathbb{P}(A|B) = 1/6$.



2 Conditional Probability

We can summarize the previous example as follows:

- What happened reduces our world to the event

$$B = \{\text{first die shows } 3\} = \{(3,1),(3,2),\dots,(3,6)\}$$

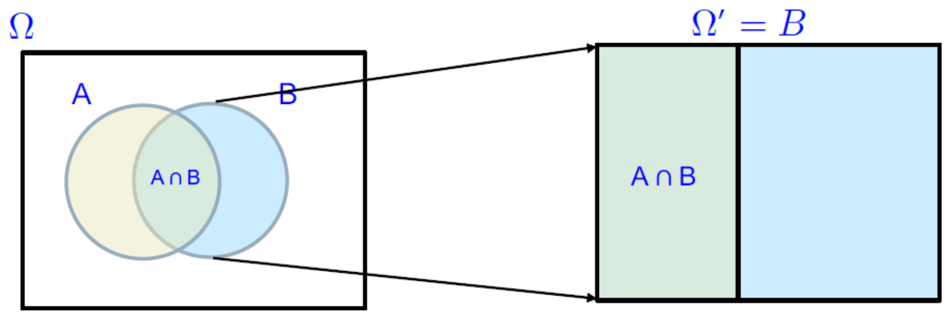
- Consider $A = \{\text{the sum is more than } 8\} = \{(3,6),(4,5),(4,6),\dots,(6,6)\}$
- Question is reformulated as "In what proportion of cases in B will also A occur?"
- "How does the probability of both A and B compare to the probability of B only?"

We now give the formal definition of conditional probability. Conditioning the original sample space means changing the perspective: instead of finding the area of A inside Ω, we are finding the area of $A \cap B$ inside B.

Definition 1 (Conditional Probability)**.** Let $\mathbb{P}(B) > 0$. The Conditional Probability of A, given B is defined as

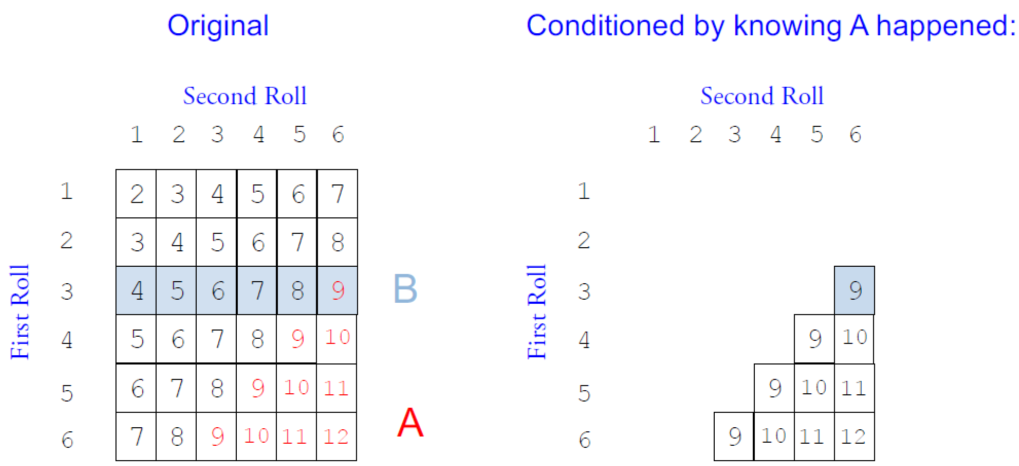
$$\mathbb{P}(A|B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}$$

Note that the conditional probability is the probability of $A \cap B$ out of the probability of B. One can check that the conditional probability $\mathbb{P}(\cdot|B)$ satisfies the nonnegativity, unit measure (over $\Omega' = B$) and additivity in disjoint events. This shows conditional probability is a well-defined probability law with the new sample space $\Omega' = B$. If $B = \Omega$, then the conditional probability recovers the original definition of probability. It shows how the observation of event B would affect the event A.



Example 2 (Toss of Two Dice)**.** Let us consider again the experiment of rolling two dices. We define A and B as before

$$A = \text{"The total dots is more than 8"}$$
$$B = \text{"The first die shows 3 dots"}$$



In Example 1 we compute $\mathbb{P}(A|B)$. Now let us use the definition of conditional probability to compute $\mathbb{P}(B|A)$, which is given as follows

$$\mathbb{P}(B|A) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(A)} = \frac{1/36}{10/36} = \frac{1}{10}.$$

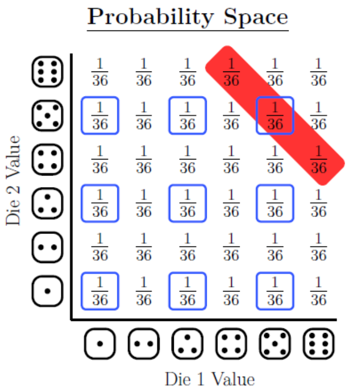
Note we have used the computation of $\mathbb{P}(A \cap B)$ and $\mathbb{P}(A)$ in Example 1. This shows that after observing the event A, the probability of B to happen is $1/10$.

Example 3 (Toss of Two Dice)**.** Let us continue to consider the experiment of rolling two dices. Now we consider different events.

Two dice have both rolled odd. What are the chances the sum is 10?

$$\mathbb{P}(\text{Sum is 10}|\text{Both are Odd}) = \frac{\mathbb{P}((\text{Sum is 10})\ \text{AND}\ (\text{Both are Odd}))}{\mathbb{P}(\text{Both are Odd})}$$

We show the calculation in the following figure



1. $$\mathbb{P}[\text{Sum is 10}] = \frac{3}{36} = \frac{1}{12}.$$
2. $$\mathbb{P}[\text{Both are Odd}] = \frac{9}{36} = \frac{1}{4}.$$
3. $$\mathbb{P}[(\text{Sum is 10})\ \text{AND}\ (\text{Both are Odd})] = \frac{1}{36}.$$
4. $$\mathbb{P}[\text{Sum is 10} \mid \text{Both are Odd}] = \frac{1}{36} \div \frac{1}{4} = \frac{1}{9}.$$

3 Understanding of Conditional Probability $\mathbb{P}(A|B)$

Suppose we repeat the same experiment several times, then the frequency of an event is the number of this event happens divided by the number of total experiments. Suppose we toss 100 coins. If there are 43 heads then the frequency for the event "H" is $43/100$ and the frequency for the event "T" is $57/100$. Frequency is an empirical quantity. It depends on the realization of experiments. A basic law in probability is that the frequency would converge to the probability if the number of experiment goes to infinity.

Proposition 1 (Frequency serves as empirical estimates of probability)**.** *Suppose we toss a coin n times and observe n_h heads. Then the frequency is*

$$\hat{p} = \frac{n_h}{n} \to \mathbb{P}(head) \text{ as } n \to \infty.$$

Therefore, frequency (an empirical quantity) serve as an approximation of probability (unknown). We can use frequency to understand the conditional probability.

- n_B outcomes in event B when you repeat an experiment n times

$$\hat{p}_B = \frac{n_B}{n} \to \mathbb{P}(B) \text{ as } n \to \infty.$$

- Of the n_B outcomes in B, the number also in A is $n_{A\cap B}$

$$\hat{p}_{A\cap B} = \frac{n_{A\cap B}}{n} \to \mathbb{P}(A \cap B) \text{ as } n \to \infty.$$

- The frequency of outcomes in A among those outcomes in B is $n_{A\cap B}/n_B$

$$\frac{n_{A\cap B}}{n_B} = \frac{n_{A\cap B}}{n} \times \frac{n}{n_B} = \frac{\hat{p}_{A\cap B}}{\hat{p}_B} \to \frac{\mathbb{P}(A\cap B)}{\mathbb{P}(B)} = \mathbb{P}(A|B)$$

- This shows that $\mathbb{P}(A|B)$ relates to the frequency of outcomes in A among those outcomes in B. This shows that the definition of conditional probability matches our rational: the definition of conditional probability in Definition 1 indeed tells us how likely the event A would happen under the constraint that B has already happened.

4 Law of Total Probability

We now introduce a result to compute the probability of an event when we are given the posterior probability.

Theorem 2 (Law of Total Probability)**.** *Let $A_1, A_2, \ldots, A_n$ be a partition of sample space Ω. Let B be any event. Then*

$$\mathbb{P}(B) = \sum_{i=1}^{n} \mathbb{P}(A_i \cap B) = \sum_{i=1}^{n} \mathbb{P}(A_i)\mathbb{P}(B|A_i).$$

Note $\mathbb{P}(B|A_i)$ are posterior probabilities. Therefore, Theorem 2 tells us how to compute the prior probability $\mathbb{P}(B)$ once we are given the posterior probabilities with respect to a partition of the sample space. Below we sketch the proof.

Proof of Theorem 2.

- Since A is a partition of the sample space, we know

$$B = (B\cap A_1) \cup (B\cap A_2) \cdots \cup (B\cap A_n).$$

Since $A_1, \ldots, A_n$ are disjoint, we know $B\cap A_1, B\cap A_2, \ldots, B\cap A_n$ are disjoint.

- By the additivity of probability we know

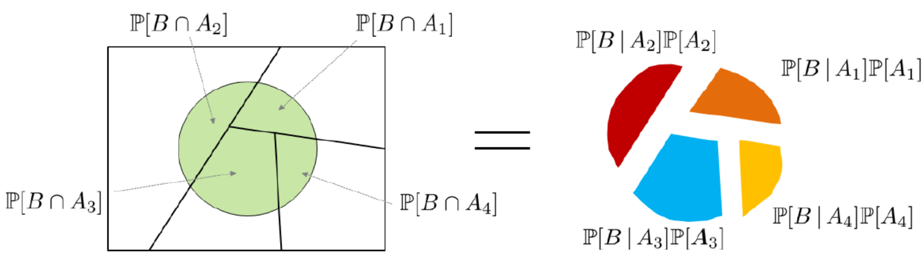
$$\mathbb{P}(B) = \sum_{i=1}^{n} \mathbb{P}(A_i \cap B)$$

- By the definition of condition probability, we know

$$\mathbb{P}(B|A_i) = \mathbb{P}(A_i\cap B)/\mathbb{P}(A_i) \Longrightarrow \mathbb{P}(A_i\cap B) = \mathbb{P}(A_i)\mathbb{P}(B|A_i)$$

□

We can summarize the proof in the following figure.



Now we give an example to show how to apply Theorem 2 in practice.

Example 4 (Probability of a Student with Long Hair)**.** We are interested in the probability of a student with long hair.

- Suppose we have a class of 60% girls and 40% boys.

- We observe that 30% of girls have long hair, and 20% of boys have long hair.
- We choose a student uniformly at random from the class.
- What is the probability that the chosen student will have long hair?

Solution 1. To answer this question, we first need to formulate it in the language of probability.

- The sample space is $\Omega = \{\text{the whole class of students}\}$, to which a partition is
$$A_1 = \{\text{the student is a girl}\}, A_2 = \{\text{the student is a boy}\}$$
- Event of interest is $B = \{\text{the student has long hair}\}$
- According to first information, we know $\mathbb{P}(A_1) = 0.6$ and $\mathbb{P}(A_2) = 0.4$
- According to the second information, we know
$$\mathbb{P}(B|A_1) = 0.3 \qquad \mathbb{P}(B|A_2) = 0.2$$
- By the law of total probability, we have
$$\mathbb{P}(B) = \mathbb{P}(A_1)\mathbb{P}(B|A_1) + \mathbb{P}(A_2)\mathbb{P}(B|A_2) = (0.6)(0.3) + (0.4)(0.2) = 0.26$$
That is, 26% chance that the randomly chosen student has long hair.

5 Product Rule

Product rule tells us how to compute the probability of an intersection of several events. According to the definition of conditional probability, we know

$$\mathbb{P}(A|B) = \mathbb{P}(A \cap B)/\mathbb{P}(B) \Longleftrightarrow \mathbb{P}(A \cap B) = \mathbb{P}(B)\mathbb{P}(A|B) \tag{5.1}$$

Applying (5.1) with $B = C \cap D$ gives

$$\mathbb{P}(A \cap C \cap D) = \mathbb{P}(C \cap D)\mathbb{P}(A|C \cap D) = \mathbb{P}(D)\mathbb{P}(C|D)\mathbb{P}(A|C \cap D).$$

We can continue this process and derive the product rule.

Theorem 3 (Product Rule)**.** *Suppose we have a sequence of events $A_1, \ldots, A_n$. The probability of the intersection is equal to*

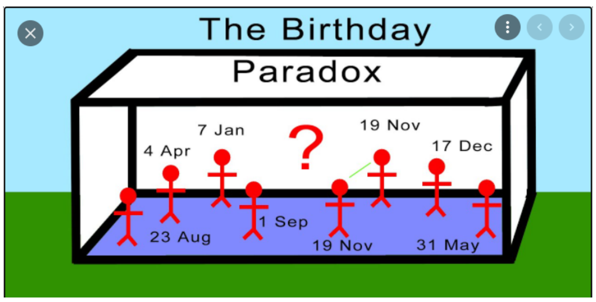
$$\mathbb{P}(A_1 \cap \cdots \cap A_n) = \mathbb{P}(A_1)\mathbb{P}(A_2|A_1)\cdots\mathbb{P}(A_n|A_1 \cap \cdots \cap A_{n-1}) \tag{5.2}$$

In other words, the product rule shows that $\mathbb{P}(A_1 \cap \cdots \cap A_n)$ can be factorized as a sequence of conditional probabilities. Notice this product rule holds for any order. For example, the following identity also holds

$$\mathbb{P}(A_1 \cap \cdots \cap A_n) = \mathbb{P}(A_n)\mathbb{P}(A_{n-1}|A_n)\cdots\mathbb{P}(A_1|A_n \cap \cdots \cap A_2).$$

As you can see, the product rule may involve more and more conditions in the conditional probability. We provide an example to show the application of product rule in solving practical problems.

Example 5 (Birthday Paradox)**.** Suppose each person has the same probability of having each day as the birthday. Out of n (≤ 365) people, what is the probability that there are no coinciding birthdays?



Solution 2. We use the product rule to solve this problem. We need to formulate the problem in the language of probability.

- Let A_i be the event that person i does not match any of the preceding people
- We are interested in

$$\mathbb{P}(A_1 \cap \cdots \cap A_n) = \mathbb{P}(A_1)\mathbb{P}(A_2|A_1)\cdots\mathbb{P}(A_n|A_1 \cap \cdots \cap A_{n-1})$$

- The product rule above requires us to compute $\mathbb{P}(A_i|A_1 \cap \cdots \cap A_{i-1})$ for each i. The condition $A_1 \cap \cdots \cap A_{i-1}$ means that the first $i-1$ people has already occupied $i-1$ positions out of 365 days. Then the i-th person only has $365-i+1$ choices to make sure his/her birthday not coinciding with any of the previous $i-1$ people. This conditional probability should be $365-i+1$ out of 365

$$\mathbb{P}(A_i|A_1 \cap \cdots \cap A_{i-1}) = \frac{365-i+1}{365}$$

- It then follows that

$$\mathbb{P}(A_1 \cap \cdots \cap A_n) = \left(\frac{365}{365}\right)\left(\frac{364}{365}\right)\cdots\left(\frac{365-n+1}{365}\right)$$

We can use the above equation to compute the probability $\mathbb{P}(A_1 \cap \cdots \cap A_n)$ for different n

88% for $n = 10$	11% for $n = 40$
59% for $n = 20$	3% for $n = 50$
29% for $n = 30$	0.00003% for $n = 100$

From the above calculations, we know that if $n = 50$ the probability of having no coinciding birthday is very low. This is a bit surprising at a first glance since 50 is not large as compared to 365. This observation is called the birthday paradox. We plot $\mathbb{P}(A_1 \cap \cdots \cap A_n)$ as a function of n in the following figure