Lecture 7: Probability

Rajat Mittal

IIT Kanpur

All of us encounter various situations in our life where we need to take a decision based on the chance/likelihood/probability of some event. We will try to make a mathematical model of these situations and see how this *chance or probability* can be quantized.

1 Definitions

Suppose we are interested in calculating/computing the chance of an outcome in a certain experiment. The experiment could be tossing a coin, throw of a die or picking a random number.

The set of all possible outcomes is known as *sample space*, it is a set and is denoted by Ω. We will be studying probability mostly in the context of use in computer science. Hence *with very high probability* our sample sets will be discrete.

Exercise 1. What is the sample space for a coin toss, sequence of coin toss, throw of a die and picking a random number.

For all these experiments, we are interested in a certain outcome or a subset of outcomes from the sample space. A subset of the sample space is known as an *event*. Our task is to model the probability of different events.

Not all subsets of the sample space need to be interesting and it might be a tedious task to define probability on every subset. But for the probability to make sense, if A, B are events, then A^C and $A \cup B$ should also be events.

This intuition gives rise to the concept of a *sigma-field*. A collection of subsets $\mathcal{F}$ of the sample space Ω is called a sigma field, if,

1. Ω is in $\mathcal{F}$.
2. Complement of a set in $\mathcal{F}$ is in $\mathcal{F}$.
3. Countable unions of sets in $\mathcal{F}$ is in $\mathcal{F}$.

Exercise 2. Show that $\mathcal{F}$ is closed under countable intersection.

With all these definitions, we are ready to define probability function. A function $P : \mathcal{F} \to [0, 1]$ is called a probability distribution function (or just probability distribution) if it satisfies,

1. $P(\Omega) = 1$.
2. If A, B are disjoint then $P(A \cup B) = P(A) + P(B)$.

Exercise 3. Show that the second rule above implies the corresponding property for countable union. Why does it stop for countable union?

Exercise 4. Prove that $P(A \cup B) = P(A) + P(B) - P(A \cap B)$.

Using the above concepts, we have modelled probability/chance/odds in an experiment. To summarize, say, we perform an experiment and are interested in the probability of various events in the experiment. The set of outcomes of the experiment will be called the sample space Ω. Any subset of Ω is an event.

We can define probability of which events are of interest, by specifying a *sigma*-field. The sigma field should be a subset of 2^{Ω} and satisfy conditions **??**.

Exercise 5. Why did we define a *sigma*-field? Why not take all the subsets of Ω as the interesting field.

A probability distribution function assigns probability to every interesting event (element of *sigma*-field) in a consistent way. By consistent, we mean that the probability function satisfies conditions **??**

Exercise 6. Suppose Amitabh (from Sholay) tosses a coin twice and is interested in finding the probability that both coins come out to be head. If coin comes head with 10% chance then what is the probability distribution function for this experiment? What is the sample space and *sigma*-field?

Let's take an example which is of interest in real life (as opposed to mathematical life). Your cousin tells you to that she has cards numbered from 1 to 1000. She will pick a card at random and if it is divisible by 2 or 5 she will pay you 100 rupees. Otherwise you will pay her 200 rupees.

Should you accept the bet. If you want to make a bet, how much money can you pay her?

Let's model the situation as a probability distribution function. Define the sample space as $\Omega = \{1, 2, \cdots, 1000\}$, set of all possible card numbers. The *sigma*-field will the set of all subsets $\mathcal{F} = 2^{\Omega}$.

We will assume that the card is picked uniformly at random, that is, the probability of obtaining a particular number in the range 1 to 1000 is 1/1000. This defines a probability distribution function for all $S \in \mathcal{F}$,

$$P(S) = \frac{|S|}{1000}.$$

Observe that we need to find the size of the set of numbers divisible by 2 or 5 and lie between 1 and 1000.

Exercise 7. Show that the numbers divisible by 2 or 5 between 1 and 1000 is 600.

The probability of you winning the game is $600/1000 = 3/5$. So *odds* of you winning are $3:2$ worse than $2:1$. So you should not accept the bet. But the bet will be favorable to you if you pay her less than 150 rupees.

Consider another example of a family. What is the probability that in a family with 5 kids, there are more girls than boys? Denote g for girls and b for boys. Then the sample space Ω is the set of all possible strings of g, b with length 5. Again, the *sigma*-field will be the set of all possible subsets of Ω. All possible strings are equally likely. Hence, we are interested in number of strings with length 5, which have more g's than b's. There are 32 possibilities, you can check that 16 of them have more girls than boys. So the probability is 1/2. This can be come up with directly by observing the symmetry between boys and girls.

Their is another way to model the same situation. The sample space will stay the same, Ω is the set of all possible strings of g, b with length 5. The difference is, σ-field is going to have only 4 elements $\{\emptyset, \Omega, A, A^C\}$.

Exercise 8. Show that for any A this is a σ-field.

Choose A to be the subset of Ω which has more girls than boys. By symmetry, the probability of A and A^C is the same. So we get that there are more girls than boys with probability 1/2.

What is the probability if there are 6 kids? If there are 6 kids then b's and g's could be equal. The number of such cases are $\frac{6!}{3!3!} = 20$. So the number of cases when girls are more than boys is $(64 - 20)/2 = 22$ and hence the probability is 22/64.

2 Conditional probability and Independence

If we pick a random person in Kanpur, there is a certain probability that this person is ill. But if we pick a random person in Kanpur from a hospital, the probability will differ.

The idea is, given two events A, B of *sigma*-field. The probability that A happens can depend upon whether B has happened or not. In the previous 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 phenomenon, we define *conditional probability.* Given two events A, B, the conditional probability of A given B is defined by,

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

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

Exercise 9. What is the probability that sum of the numbers on two throws is 3 given that the sum is a prime?

A related question is of independence between events, consider the following two questions.

1. What is the probability of obtaining two heads 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?

It is clear that if Euler misses the school on first day, he might miss it the next day too with high probability (because he might be out of station or ill). In some 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.

Two events A, B are said to be independent if,

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

Exercise 10. What is the relationship between independence and conditional probability of two events?

Let's look at the application of these concepts. There was a survey conducted by the Health dept. in a hospital (with Asthma and Diabetes patients), it found that people who had diabetes did not have Asthma with higher probability as compared to the general population. This suggests that people who have Asthma have less likelihood of getting Diabetes.

It turns out that even if having Asthma and Diabetes are independent of each other, there will be a negative relation between Asthma and Diabetic patients in a hospital.

In other words, suppose A, B are two independent events. They will not be mutually independent if we consider the conditional probabilities given $A \cup B$. That is, events $\frac{A}{A \cup B}$ and $\frac{B}{A \cup B}$ will be negatively related even if A, B are independent. This is known as *Berkson's Paradox*.

To make it more quantitative, consider a sample of 1000 balls. We know that 100 of them are red, 50 of them are shiny and 5 of them are red and shiny. The probability of being shiny is $1/20$ and also the probability of a red ball being shiny is $5/100 = 1/20$. Hence being red and being shiny are independent.

Say, we pick only the balls which are red and shiny. Then the probability that a ball is shiny is $1/3$ but a red ball being shiny remains at $1/20$. This will show that the a red ball is mostly not shiny.

Exercise 11. Say A be the event that the ball is shiny and B being the event that the ball is red. Prove that $Pr\left(\frac{A}{A \cup B}\right) > Pr(A)$. Convince yourself that this is the reason why events A and B seem to be negatively related.

Exercise 12. Convince yourself that same thing happened when health department surveyed in a hospital in the above example.

3 Bayes theorem

This is an example taken from Daniel Kahneman's book [**?**].

Exercise 13. 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.

Such examples are very common in everyday life. Let's look at other example. Suppose a scientific theory predicts that there will be a solar eclipse on 1st Oct., 2015 with high probability. If we observe that there is a solar eclipse on 1st Oct, what is the probability that the theory is correct.

Such problems are called hypothesis testing. Let event A be that the scientific theory is true and B be the event that solar eclipse happens on 1st Oct.. So we know the conditional probability $P\left(\frac{B}{A}\right)$ and are interested in the conditional probability $P\left(\frac{A}{B}\right)$.

Example 1. Frame the problem about Ramanujan's department above in terms of hypothesis testing.

Bayes' theorem gives answer to such problems with a very clean expression.

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

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

Note 2. The denominator is mostly obtained using the formula $P(B) = P\left(\frac{B}{A}\right)Pr(A) + P\left(\frac{B}{A^c}\right)Pr(A^c)$.

The proof is given as an exercise in the assignment.

Let's take some examples and see how Bayes' theorem can be applied in various settings.

– Suppose there is a test for early detection of cancer and a study shows that it is very successful. If a person has cancer then the test diagnoses it with probability .9. 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?

  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. Let's try to solve it using Bayes' theorem. Say A be the event that person has cancer and B be the event that test outputs that person has cancer. Then,

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

  We realize that the information given to us is incomplete. It at least needs the ratio between $P(A)$ and $P(A^c)$ (this is known as *base rate*). Say 1% of the general population has cancer. Then,

  $$Pr\left(\frac{A}{B}\right) = \frac{.9 \times 1}{.9 \times 1 + .1 \times 99} \approx .1.$$

  This shows that base rate matters a lot in this calculation and should not be ignored.

– 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.

  Exercise 14. You are told that recently there was an accident involving a taxi, what is the probability that the taxi was white?

3.1 Monty Hall problem

This famous problem is posed in the context of a game show. There are 3 doors and behind one of them has a car and other two have goats hidden behind them. You are asked to pick a door, then the game show host (Monty Hall) opens one of the other two doors and reveals a goat. Assuming that you are not interested in a goat, should you switch the door?

This problem is very famous because of the counter-intuitive nature of the result. We can calculate the exact probability of whether switching helps or not. We will 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 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 $P\left(\frac{D_1}{B}\right)$.

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

Note 3. 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 15. Convince yourself that the formula is correct.

We know that $Pr\left(\frac{B}{D_2}\right)$ 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.

4 Assignment

Exercise 16. For the events $A_1, A_2, \cdots, A_n$, prove,

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

Exercise 17. How many numbers are there between 1 and 1000 divisible by 2,3 or 5?

Exercise 18. 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 19. Read more about Monty Hall problem and its variations from Wikipedia.

Exercise 20. Give the proof of Bayes' Theorem.

Exercise 21. 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?

References

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