Week 2: Conditional Probability and Bayes formula

We ask the following question: suppose we know that a certain event B has occurred. How does this impact the probability of some other A. This question is addressed by conditional probabilities. We write

$$P(A|B) = \textbf{the conditional probability of } A \text{ given } B$$

Example: Suppose a family has two children and suppose one of the children is a boy. What is the probability that both children are boys?

To answer this question we suppose that it is equally likely to have boys or girls. The the sample space for the children is $S = \{BB, BG, GB, GG\}$ where for example BG means that the first child is a boy and the second is a girl. We have

$$P(BB) = P(BG) = P(GB) = P(GG) = \frac{1}{4}$$

Let us consider the events

$$A = \{BG, GB, BB\}, \text{ one of the children is a boy}$$

and

$$B = \{BB\}, \text{ both children are boys}$$

Since all events are equally likely we if we know that F has occurred, we assign now the new probabilities $1/3$ to all three events in F and thus we obtain

$$P(A|B) = \frac{1}{3}$$

■

Definition of conditional probability: Given an event B, we assign new probabilities for each outcome in the sample space $p(i|B)$. Since we know that B has occurred then we must have

$$p(i|B) = 0, \quad \text{for } i \in \overline{B}$$

and if $i \in B$ we require that the probabilities $p(i|B)$ have the same relative magnitude than $p(i)$. That is we require that for some constant c

$$p(i|B) = cp(i) \quad \text{for } i \in B$$

But we must have
$$1 = \sum_{i \in B} p(i|B) = c \sum_{i \in B} p(i) = cP(B)$$
and thus
$$c = \frac{1}{P(B)}$$
So we define

Condititional probability of i given B: $\qquad p(i|B) = \dfrac{p(i)}{P(B)}, \quad i \in B$

If we consider an event A then we have
$$P(A|B) = \sum_{i \in A} P(i|B) = \sum_{i \in A \cap B} \frac{p(i)}{P(B)} = \frac{P(A \cap B)}{P(B)}$$
and so obtain

Condititional probability of A given B: $\qquad P(A|B) = \dfrac{P(A \cap B)}{P(B)}$

It is also useful to think of this formula in a different way: we can write

---

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

---

that is to compute the probability that both A and B occurs can be computed as the probability that B occurs time the conditional probability that A occurs given B.

Finally we give one more application of this formula: Suppose you want to compute the probability of an event F. Sometimes it is much easier to compute $P(F|E)$ or $P(F|\overline{E})$. We write

$$P(A) = P(A \cap B) + P(A \cap \overline{B}) = P(A|B)P(B) + P\left(A|\overline{B}\right)P\left(\overline{B}\right)$$

This formula is often very useful if one chooses B in a smart way.

$$\boxed{\textbf{Conditioning}: \ P(A) \ = \ P(A|B)P(B) + P(A|\overline{B})P(\overline{B})}$$

More generally we can condition on a collection of n events provided they are pairwise disjoint and add up to all the sample space. If $S = B_1 \cup B_2 \cup \cdots B_n$ and the B_i are pairwise disjoint $B_i \cap B_j = \emptyset$, $i \neq j$ then we have

$$\boxed{\textbf{Conditioning}: \ P(A) \ = \ P(A|B_1)P(B_1) + P(A|B_2)P(B_2) + \cdots + P(A|B_n)P(B_n)}$$

Example: The game of crap The game of crap is played as follows:

- Roll two dice and add the numbers obtained.

- If the total is 7 or 11 you win.

- If the total is 2, 3, 12 you loose.

- If the total is any other number (i.e., 4, 5, 6, 8, 9, 10) then this number is called "the point".

- Roll the pair of dice repeatedly until you obtain either "the point" or a 7.

- If you roll a 7 first you loose.

- If you roll "the point" first you win.

We compute the probability to win at this game. To do this we condition on the following events "first roll is 7 or 11", "the first roll is 2, 3, 12" "the point is 4", "the point is 5", etc.... We have then

$$
\begin{aligned}
P(\text{Win}) \ = \ & P(\text{Win}|\text{first roll 7 or 11})P(\text{first roll 7 or 11}) \\
& + P(\text{Win}|\text{first roll 2 3 or 12})P(\text{first roll 2, 3, or 12}) \\
& + \sum_{i \in \{4,5,6,8,9,10\}} P(\text{Win}|\text{point is } i)P(\text{point is } i)
\end{aligned}
$$

Most of these probabilities are easy to compute: The only one which requires some thought is $P(\text{Win}|\text{point is } i)$. Take for example the point to be 4. To compute this probability we argue that we roll the dice until we get a 4 or a 7 at which point the game

stop. It does not matter how many times we roll the dice, the only thing which matters is that to win we need a 4 rather than a 7. So

$$P(\text{Win}|\text{point is 4}) = P(\text{roll a 4}|\text{roll a 4 or a 7}) = \frac{\frac{3}{36}}{\frac{3}{36} + \frac{6}{36}} = \frac{3}{9}$$

We leave it to the reader to verify that

$$P(\text{Win}) = 1 \times \frac{8}{36} + 0 \times \frac{4}{36} + \frac{3}{9} \times \frac{3}{36} + \frac{4}{10} \times \frac{4}{36} + \frac{5}{11} \times \frac{5}{36} + \frac{5}{11} \times \frac{5}{36} + \frac{4}{10} \times \frac{4}{36} + \frac{3}{9} \times \frac{3}{36} = .49293...$$

which shows that Crap was surely designed by a someone with a knowledge of probability...

∎

Example: The game of roulette: We consider several variants of roulette.

- Las Vegas roulette has 38 numbers, 0 and 00 which are green and 1 to 36, half them being red and half of them being black. If you bet on red the probability to loose is $20/38 = 0.526315789$.

- Monte-Carlo roulette (1st version) has 37 numbers, 0 and 1 to 36, half them being red and half of them being black. If you roll a 0 then you are sent to prison (P1). At the next spin if you get a red you get your bet back (and nothing more), if you get black or 0, you loose. The probability to win is $18/37$, and the probability to loose is obtained by conditioning on the first spin

$$
\begin{aligned}
P(\text{ lose }) &= P(\text{ lose } | \text{ black })P(\text{ black }) + P(\text{ lose } | P1)P(P1) \\
&= 1 \times \frac{18}{37} + \frac{19}{37} \times \frac{1}{37} = 0.50036523
\end{aligned}
\tag{1}
$$

- Monte-Carlo roulette (2st version) is played as in the 1st version but with a second prison (P2). If you are in the first prison P1 you loose if the next spin is black and if the next spin is a 0 you are sent to the second prison P2. In P2 you loose if you get a black or 0 and are sent back to P1 if you get a red. The probability to lose is obtained by conditioning several times. First we have

$$P(\text{ lose }) = 1 \times \frac{18}{37} + P(\text{ lose } | P1)\frac{1}{37}$$

But if we are in P1 and condition on the next spin we have

$$P(\text{ lose } | P1) = \frac{18}{37} + P(\text{ lose } | P2)\frac{1}{37}$$

4

and similarly if we are in P2

$$P(\text{ lose } | P2) = \frac{19}{37} + P(\text{ lose } | P1)\frac{18}{37}$$

The last two equations can be combined to find the value of $P(\text{ lose } | P1)$

$$P(\text{ lose } | P1) = \frac{18}{37} + \frac{1}{37}\left(\frac{19}{37} + P(\text{ lose } | P1)\frac{18}{37}\right)$$

which gives

$$P(\text{ lose } | P1) = \frac{685}{1351}$$

and so

$$P(\text{ lose }) = \frac{18}{37} + \frac{1}{37}\frac{685}{1351} = 0.50019004.$$

■

Conditional probability and independence: It is natural to define independence between two events in terms of conditional probabilities. We will say that A is *independent* of B if the probability that A occurs does not depend on whether B has occurred or not. In other words

$$A \textbf{ independent of } B \quad \text{if } P(A|B) = P(A)$$

Now using the definition of conditional probability this is equivalent to

$$\frac{P(A \cap B)}{P(B)} = P(A) \quad \text{or } P(A \cap B) = P(A)P(B)$$

The formula on the right is symmetric in A and B and so if A is independent of B then B is also independent of A. So we have

$$A \text{ and } B \text{ are } \textbf{independent} \quad \text{if } P(A \cap B) = P(A)P(B)$$

Bayes formula: A particular important application of conditional probability is Bayes formula. At the basic mathematical level it is a formula which relates $P(A|B)$ and $PB|A)$. It is very easy to derive but its importance is hard to overemphasize. We have

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

from which we conclude that

$$\boxed{\textbf{Bayes Formula} \quad P(A|B) = \frac{P(B|A)P(A)}{P(B)}}$$

One should interpret this formula as follows: before we do an experiment (given by the event B) the probability of A is $p(A)$. But after the experiment the probability that A occurs is $P(A|B)$. So Bayes formula is a way to understand how we learn about the world if the world is uncertain. We perform experiments and obtain so knowledge which changes the probabilities. This suggest the following terminology

$P(A)$ is he **prior probability**

$P(A|B)$ is the **posterior probability**

B the **evidence**

Sometimes one finds another version for Bayes formula where the denominator $P(B)$ is written using conditioning:

$$P(A|B) = \frac{P(B|A)P(A)}{P(B|A)P(A) + P(B|\bar{A})P(\bar{A})}$$

Example: Witness reliability: Often question arise which are expressed directly in term of conditional probabilities in which case Bayes formula is very handy. Imagine the following example: after a robbery the thief jumped into a taxi and disappeared. An eyewitness on the crime scene is telling the police that the cab is yellow. In order to make sure that this testimony is worth something the assistant DA makes a Bayesian analysis of the situation. After some research he comes up with the following information:

- In that particular city 80% of taxis are black and %20 of taxis are yellow.

- Eyewitness are not always reliable and from past experience it is expected that an eyewitness is %80 accurate. He will identify the color of a taxi accurately (yellow or black) 8 out 10 times.

Equipped with this information assistant DA first defines adequate events

- $true = Y$ means that the color of the taxi was actually yellow while $true = B$ means that it was black.

- *report* $= Y$ means that the eyewitness identified the color of the taxi as actually yellow while *report* $= B$ means that it was reported as black.

The goals is to compute $P(true = Y|report = Y)$ and using Bayes formula he finds

$$P(true = Y|report = Y) = \frac{P(report = Y|true = Y)P(true = Y)}{P(report = Y)}$$

The he notes that $P(true = Y) = .2$ (this is the prior probability) and $P(report = Y|true = Y) = .8$ (this is the accuracy of witness testimony). Finally to compute $P(report = Y)$ he argues that this depends on the actual color of the taxi so using conditional probability he finds

$$
\begin{aligned}
P(report = Y) &= P(report = Y|true = Y)P(true = Y) \\
&\quad + P(report = Y|true = B)P(true = B) \\
&= .8 \cdot .2 + .2 \cdot .8 = .032
\end{aligned}
\tag{2}
$$

Putting everything together we finds

$$P(true = Y|report = Y) = \frac{P(report = Y|true = Y)P(true = Y)}{P(report = Y)} = \frac{.16}{.32} = \frac{1}{2}$$

and so the eyewitness testimony does not provide much useful certainty. ∎

Example: Spam filter A situation where Bayesian analysis is routinely used is your spam filter in your mail server. The message is scrutinized for the appearance of key words which make it likely that the message is spam. Let us describe how one one of these filters might work. We imagine that the evidence for spam is that the subject message of the meal contains the sentence "*check this out*". We define events

- *spam* which means the message is spam.

- "*check this out*" which means the subject line contains this sentence.

and we are trying to compute the conditional probability

$$P(spam|"check this out")$$

In order to compute this probability we need some information and note that from previous experience

- 40% of emails are spam

- 1% of spam email have "*check this out*" in the subject line while % .4 of non-spam emails have this sentence in the subject line.

Using Bayes formula we find

$$P(spam|"check\ this\ out") = \frac{P("check\ this\ out"|spam)P(spam)}{P("check\ this\ out")}$$

Now we have $P(spam) = .4$ while $P("check\ this\ out"|spam) = .01$. To compute $P("check\ this\ out")$ we condition and find

$$
\begin{aligned}
P("check\ this\ out") &= P("check\ this\ out"|spam)P(spam) \\
&+ P("check\ this\ out"|not\ spam)P(not\ spam) \\
&= .01 \cdot .4 + .004 \cdot .6 = .0064
\end{aligned}
$$

and so we find

$$P(spam|"check\ this\ out") = \frac{.004}{.0064} . = \frac{5}{8} = .625$$

In this case it is a (weak signal) that the message is spam and further evidence is required to weed out the message. ∎

Example: the Monty's Hall problem At a game show the host hides a prize (say $ 1 million) behind one of three doors and nothing of much value behind the two remaining doors (in the usual story two goats). The contestant picks one of three doors, let us say say door 1, and then the game show host opens one of the remaining door, let us say he opens door 3 which reveals a goat. The contestant is then given the choice to either switch to door 2 or keep door 1. What should he do?

We will argue that he should switch to door 2 since there is a greater probability to find the prize behind door 2 than behind door 1. The trick in this problem is to carefully make your assumptions precise and we assume, quite reasonably, that

- The $1 million is put randomly behind any door, that is, the contestant, upon choosing a door, has probability 1/3 to find the prize.

- The host show knows behind which doors the prize is and always opens an empty door. If he has two empty doors he can open then he chooses one of the two doors at random.

There are many ways to find the solution and we will present two solution, one which involves just an elementary consideration and the second which uses Bayes formula.

Solution 1: To assign probabilities properly we name the two goats G1 and G2 and note that the following arrangements are possible

Door 1	Door 2	Door 3
P	G1	G2
P	G2	G1
G1	P	G2
G2	P	G1
G1	G2	P
G2	G1	P

Since the prize are arranged randomly we may assume that all the 6 arrangements have equal probabilities $1/6$. Now we assume that the contestant opens door 1. In the first case the contestant has the prize behind his door and the host will open either door 2 or door 3 in which case the contestant will lose if he switches door. The second case is similar. In the third case the hist will open door 3 and the contestant will win if he switches door. The fourth fifth and sixth case are similar. So in 2 out of 6 case you win by keeping your door while in 4 out of 6 cases you win by switching. Hence the probability to win if you switch is $2/3$. So you should switch.

Solution 2: Let us start to analyze this problem when the the contestant has chosen door 1 and the host has opened door 3. The probability for the contestant to keep his door and win is a conditional probability

$$P(keep\ and\ win\ ) = P(\ prize\ door\ 1|\ host\ door\ 3)$$

with obvious notations. To compute this we use Bayes formula and obtain

$$P(\ prize\ door\ 1|\ host\ door\ 3) = \frac{P(\ host\ door\ 3|\ prize\ door\ 1)P(\ prize\ door\ 1)}{P(\ host\ door\ 3)}$$

By our first assumption we have

$$P(\ prize\ door\ 1) = \frac{1}{3}$$

while our second assumption implies that

$$P(\ host\ door\ 3|\ prize\ door\ 1) = \frac{1}{2}$$

since the host can choose between door 2 or door 3. To compute $P(\ host\ door\ 3)$ we condition on the location of the prize and we have

$$
\begin{aligned}
P(\ host\ door\ 3) &= P(\ host\ door\ 3|\ prize\ door\ 1)P(\ prize\ door\ 1) \\
&+ P(\ host\ door\ 3|\ prize\ door\ 2)P(\ prize\ door\ 2) \\
&+ P(\ host\ door\ 3|\ prize\ door\ 3)P(\ prize\ door\ 3) \\
&= \frac{1}{2} \cdot \frac{1}{3} + 1 \cdot \frac{1}{3} + 0 \cdot \frac{1}{3} \quad (3)
\end{aligned}
$$

9

Putting everything together we find

$$P(\textit{keep and win }) = P(\textit{ prize door } 1| \textit{ host door } 3) = \frac{\frac{1}{3} \cdot \frac{1}{2}}{\frac{1}{2} \cdot \frac{1}{3} + 1 \cdot \frac{1}{3}} = \frac{1}{3}$$

as before. ∎

Exercise 1: A coin is tossed three times. What is the probability that two heads occur given that

- The first outcome was a head.

- The first outcome was a tail.

- The first two outcomes were heads.

- The first tow outcomes were tails.

- The first outcome was a head and the third was a head.

Exercise 2: Imagine a game where a player is handed two cards (say at the beginning of Texas Hold' Em Poker). A particularly lousy opponent actually reveals one of his two cards to you and it is an ace. Consider the following cases:

1. His first card is an ace. What is the probability he has two aces? That is compute

$$P(\textit{ 2 aces } | \textit{ first card is ace })$$

2. One of is his card is an ace. What is the probability he has two aces? That is compute
$$P(\textit{ 2 aces } | \textit{ one card is an ace })$$

3. One of his card is the ace of spade. hat is the probability he has two aces? That is compute
$$P(\textit{ 2 aces } | \textit{ one card is the ace of spade})$$

   that his second card is an ace.

4. Are you surpassed by the results?

Exercise 3: Suppose Math 478 has two section. In section I there is 12 female and 18 male students. In section II there are 20 female and 15 male students. The professor picks a section at random and then picks a student at random in that section. Compute

1. Probability that the student chosen is a female.

2. The conditional probability that the student is in section II given that she is a female

Exercise 4: False positives, Sensitivity and Specificity. Bayes formula is very much used in epidemiology. Suppose we deal with a disease and we have test for the disease. We know

- The *sensitivity* of the test is 99%, this means that if you have the disease, the test is positive with probability 0.99.

- The *specificity* of the test is 98%, this means that if you do not have the disease, the test is negative with probability 0.98.

- The *prevalence* of the disease is one in two hundred, this means that the probability to carry the disease is 0.005.

The probability that someone with a positive test actually has the disease is called the *positive predictive value* of the test. The probability that someone with a negative test actually does not have the disease is called the *negative predictive value* of the test.

Express the positive and negative predictive value of the test using conditional probabilities and compute them using Bayes formula.

Exercise 5: Monty's Hall In this problem you will analyze how important the assumptions are in Monty's Hall problem.

1. Assume as before that the prize is randomly put behind one the three doors. However after watching the game many times you notice that the show host who is usually standing near door 1 is lazy and he tends to open the door closest to him. For example if you pick door 1, he opens door 2 75% of the time and door 3 only 25% of the time. Does this modify your probability of winning if you switch (or not switch)?

2. Suppose you have been watching Monty's Hall game for a very long time and have observed that the prize is behind door 1 45% of the time, behind door 2 40% of the time and behind door 3 15% of the time. The rest is as before and you assume that the host opens door at random (if he has a choice).

   (a) When playing on the show you pick door 1 again and the host opens one empty door. Should you switch? Compute the various cases.

   (b) If you know you are going to be offered a switch would it be better to pick another door? Explain.