Chapter 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

<div style="border:1px solid">

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

</div>

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

<div style="border:1px solid">

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

</div>

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

<div style="border:1px solid">

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

</div>

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.

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

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

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

Example: The game of craps The game of craps 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 then find

$$P(\text{Win}) = \underbrace{0 \times \frac{1}{36}}_{2} + \underbrace{0 \times \frac{2}{36}}_{3} + \underbrace{\frac{3}{9} \times \frac{3}{36}}_{4} + \underbrace{\frac{4}{10} \times \frac{4}{36}}_{5} + \underbrace{\frac{5}{11} \times \frac{5}{36}}_{6} + \underbrace{1 \times \frac{6}{36}}_{7}$$
$$+ \underbrace{\frac{5}{11} \times \frac{5}{36}}_{8} + \underbrace{\frac{4}{10} \times \frac{4}{36}}_{9} + \underbrace{\frac{3}{9} \times \frac{3}{36}}_{10} + \underbrace{1 \times \frac{2}{36}}_{11} + \underbrace{0 \times \frac{1}{36}}_{12}$$
$$= \frac{244}{495} = 0.49292929...$$

which shows that craps 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 $20/38 = 0.526315789$ and so on average you gain $1 \times \frac{18}{38} - 1\frac{20}{38} = -\frac{2}{38} = .052....$ that is you lose, on average, 5.2 cents on every dollar you bet.

- 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 $\frac{18}{37} = \frac{666}{1369} = 04864...$, and the probability to loose is obtained by conditioning on the first spin

$$P(\text{ lose }) = P(\text{ lose } | \text{ black })P(\text{ black }) + P(\text{ lose } | P1)P(P1)$$
$$= 1 \times \frac{18}{37} + \frac{19}{37} \times \frac{1}{37} = \frac{685}{1369} = 0.5000365....$$

and thus the probability to tie is $\frac{18}{37} \times \frac{1}{37} = \frac{18}{1369} = 0.0131....$

In that version your average gain is $1 \times \frac{666}{1369} + 0 \times \frac{18}{1369} - 1\frac{685}{1369} = -0.01387$, that you lose 1.3 cents on very euro you bet. This suggests that la Côte-D'azur might be beat the Nevada desert from this point of view at least.

- 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

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

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

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

∎

Example: The game of blackjack Blackjack, a very popular game, has the following rules. It is played with one or several regular decks of 52 cards. The cards values are the face values for the cards 2 to 9, 10 for 10, jack, queen, or king, and for the ace the value is 1 or 11, whichever is most advantageous. The goal is to get as close as possible to a total of 21 but exceeding it (a "bust")

- The player(s) get two cards which are put face up on the table.

- The house get two cards, one face up and one face down.

- If the player or the house has gotten a total of 21 (a "blackjack") then the game ends immediately and the one with the blackjack wins.

5

- Otherwise the player has several options to choose from.

  - Hit: ask for another card. This can be done several times.
  - Stand: do nothing.
  - Double down: which means ask for exactly one extra card and then stand.
  - Split: Split your card, ask for two cards and now play two games.
  - Surrender: give your card away and get half of your bet back.

- After the player has chosen his options, the house will hit until it reaches 17 or higher and then stand or bust if the total exceeds 21.

As a warm-up let us compute the probability to get a blackjack with two cards. If there is one deck it is
$$\frac{\binom{4}{1}\binom{16}{1}}{\binom{52}{2}} = \frac{128}{52 \times 51} = 0.0483$$
since there are four aces (with a value of 11) and 16 cards with a values of 10. If you play with n decks of cards it is
$$\frac{\binom{4n}{1}\binom{16n}{1}}{\binom{52n}{2}} = \frac{128n^2}{52n(52n-1)} = \frac{128}{52(52 - \frac{1}{n})}$$

As n goes to ∞ this probability varies a little bit and decreases to 0.04737.

Now as an example suppose we have the following hands

$$\underbrace{10\heartsuit + 8\diamondsuit}_{\text{player}} \qquad \text{vs} \qquad \underbrace{J\spadesuit + ?}_{\text{house}}$$

In that case should the player stand or hit? To decide we can compute the probability that the player will not bust if he hits. For one deck this probability is $12/49 = 0.2448$ since he needs one ace, 2, or, 3. If there are n decks this probability is $12n/52n - 3 = 12/(52-3/n)$ which for large n is close to $12/52 = 0.2307$. Again we see that probability change very little when going from 1 to many decks. In any case the strategy seems fairly clear and the player should stand.

Finally let us assume that the player stand and keeps his total of 18. We now would like to compute the probability that the players wins. To do this requires a lot computations, and we just do one case to illustrate how it works. We compute the probability that the house get a total of 19, that is the conditional probability that the house get a 19 with an up card of 10. The total of 19 be achieved in many ways: for example the face down card could be a 9. If the face down card is not a 9, it could be a lower value. It cannot be a 7 or a 8 because then the house would reach 17 and 18 and stop. So it could be a 6 and the house hits. It can reach 19 by getting a 3 but again it cannot reach 19 by getting a

6, a 2 and a1, since he would reach 18 first and stop. Finally one should remember that a 1 cannot be the face down since it would have been a blackjack. A little thought shows that the house can get a 19 after the following sequences of cards

$$9, \quad 6-3, \quad 5-4, \quad 5-1-3, \quad 4-5, \quad 4-2-3, \quad 4-1-4, \quad 4-1-1-3$$

$$3-6, \quad 3-3-3, \quad 3-2-4, \quad 3-2-1-3, \quad 3-1-5, \quad 3-1-2-3, \quad 3-1-1-4, \quad 3-1-1-1-3$$

$$2-7, \quad 2-4-3, \quad 2-3-4, \quad 2-3-1-3, \quad 2-2-5, \quad 2-2-2-3, \quad ,2-2-1-4, \quad 2-2-1-1-3$$

$$2-1-6, 2-1-3-3, 2-1-2-4, 2-1-2-1-3, 2-1-1-5, 2-1-1-2-3, 2-1-1-1-4, 2-1-1-1-1-3$$

While this might seem a mess there is lot of recursive structure that you should explore. To compute (or rather estimate) the probability to get a 19 we use an infinite deck assumptions so that the probability to get any given value is simply $1/13$, irrespective of all other cards already drawn and we find

$$P(\text{dealer19}|\text{up card10}) = (1/13)+5(1/13)^2+10(1/13)^3+10(1/3)^4+5(1/3)^5+(1/3)^6 = 0.1114$$

In a very similar way one can compute that

$$P(\text{dealer 20}|\text{up card10}) = 0.3421, \quad P(\text{dealer21}|\text{up card10}) = 0.0345,$$

so that the probability that the house wins is about 0.48.

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

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

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

$$\boxed{\begin{aligned} &P(A) \text{ is he } \textbf{prior probability} \\ &P(A|B) \text{ is the } \textbf{posterior probability} \\ &B \text{ the } \textbf{evidence} \end{aligned}}$$

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

$$
\begin{aligned}
P(report = Y) &= P(report = Y | true = Y)P(true = Y) \\
&\quad + P(report = Y | true = B)P(true = B) \\
&= .8 \cdot .2 + .2 \cdot .8 = .032
\end{aligned}
\tag{1}
$$

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

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

and so we find

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

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 $P(\text{ host door } 3)$ we condition on the location of the prize and we have

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

Putting everything together we find

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

as before. ∎