Conditional Probability

Instructor: Karl Schmedders

August 2021

1 Conditional Probabilities

We start this discussion with an example — let us consider the probability that a person knowing to speak Mandarin is 30%. So if we consider the sample space of the world population, 3 out of every 10 person would know to speak Mandarin.

Now, if we gain some more information and realize that the person concerned is Indian then our probability would change drastically because not many Indians know Mandarin. Assuming that only 5% of Indians speak Mandarin, our probability has decreased from 30% to only 5%. This is an example of conditional probability.

Mathematically, Let A be the event that a person knew Mandarin and B be the event that a person is from India. In this case, $P(A) = 0.3$ and $P(A|B) = 0.5$ which is the probability that a person knows Mandarin given that he/she is from India. $P(A|B)$ can be calculated using $\frac{P(A \cap B)}{P(B)}$ where $P(A \cap B)$ is the probability that a person knows Mandarin **and** belongs to India.

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

Notice that when calculating $P(A)$ our sample space was the entire world population but in the second case, our sample space is reduced to only the set of people belonging to India. Going back to the fundamental definition of probability, we observe that in the smaller sample space the outcomes favourable to $P(A|B)$ are in the set $A \cap B$ and as mentioned before the sample space is B so we normalize the value of $P(A \cap B)$ over the value of $P(B)$ to get $P(A|B)$. See fig. 1.

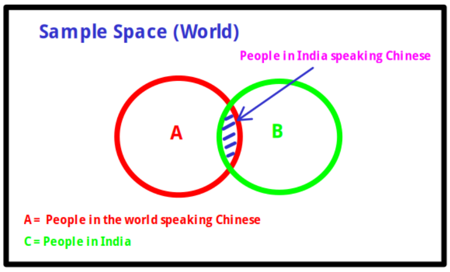

Figure 1: Venn Diagram depicting Conditional Probability

2 Dependence and Independence

The condition of the event B has changed the probability of A i.e. a change from 30% to 3%. This shows that the likelihood of event A occurring given event B has occurred is dependent on the occurrence of B, thus A and B are dependent events.

On the contrary, if the probability of A remains the same even after the occurrence of the new event B, then the events are independent.

Mathematically, if event A and B are independent then $P(A|B) = P(A)$ and obviously, $P(B|A) = P(B)$. Note that we can write (1) as:

$$P(A \cap B) = P(A|B) \times P(B) \tag{2}$$

So if A and B in (2) are independent events then, it leads to the *multiplication rule* given by:

$$P(A \cap B) = P(A) \times P(B) \tag{3}$$

This brings us to an interesting example provided in the course — **the Bathtub curve**.

First let us consider a scenario, you work at a factory and based on past empirical data, it has been calculated that a machine X's chance of failure is 10%. So it fails one day in every 10 days and works fine on 9 out of 10 days. Now, our task is to calculate the probability that it works fine two days in a row.

Let us assume W_i is the event that it works fine on a day i, then W_{i+1} is the event that it works fine on the very next day. We are supposed to find $P(W_i \cap W_{i+1})$, we know by our previous discussion that we can find this probability using (2) as:

$$P(W_i \cap W_{i+1}) = P(W_i|W_{i+1}) \times P(W_{i+1}) \tag{4}$$

But it is not possible to determine the answer using (4) as we don't know the value of $P(W_i|W_{i+1})$. It is also interesting that this value can be determined easily if we consider the events W_i and W_{i+1} to be independent but we don't have any evidence to do so (yet?).

$$P(W_i \cap W_{i+1}) = P(W_i) \times P(W_{i+1}) \tag{5}$$

This finally brings us to the bathtub curve that we have been foreshadowing, the bathtub curve is a curve used in reliability engineering as a three-part hazard function where the failure rate of a machine is plotted on the y-axis while time runs on the x-axis. The curve is shown in figure 2.

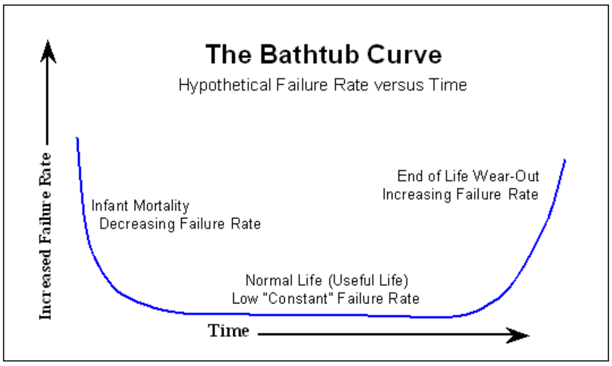

Figure 2: The Bathtub curve. *Source: weibull.com*

In layman terms, the curve tells us that when a machine is newly installed there is high chance that the machine can malfunction due to manufacturing defect or bad calibration, etc., if the machine did not fail during that period then it is likely to work for a period in which it may only fail due to random errors. After the machine has been functioning for a long time, it will eventually undergo a lot of wear and tear, and finally succumb to old age.

Now in the context of our quest for determining $P(W_i \cap W_{i+1})$, if our machine behaves according to the bathtub curve, then it is evident that the events W_i and W_{i+1} are dependent if the consider the early or the final part of the plot. We can also observe that if we consider only the middle part of the curve then the two events indeed seems independent so we can use (5) to arrive at a result.

3 Probability Tables

	Label: $P(A)$	Label: $P(A^c)$	
Label: $P(B)$	$P(A \cap B)$	$P(A^c \cap B)$	$P(B)$
Label: $P(B^c)$	$P(A \cap B^c)$	$P(A^c \cap B^c)$	$P(B^c)$
	$P(A)$	$P(A^c)$	$P(U) = 1$

Table 1: The format of a simple probability table

When we try to solve a probability problem, it is easy to keep track of probability values when we only have to consider a few events. As the number of events become increase, it becomes more difficult to calculate and recall probability values. In such a scenario we can make use of a probability table (such as the one shown in table 1). With the help of such a table, it is quite easy to calculate a conditional probability, say $P(A|B^c)$ as the required values of $P(A \cap B^c)$ and $P(B^c)$ can be readily retrieved from the table.

Note: In order to prepare a probability table for a given data, it is necessary that the events specified in the row and column headings are *mutually exclusive* and *collectively exhaustive.*

	Passenger Type				
	First Class	Second Class	Third Class	Crew	Total
Survived	203	118	178	212	711
Died	112	167	528	673	1490
Total	325	285	706	885	2201

Table 2: A table of data relevant to data about the ship *Titanic*

But in order to prepare a probability table, we have to begin from a table such as Table 2 that provides the statistics of the survival of the ship *Titanic* grouped by the passenger type. In this table, each cell provides unique information depending on the column and row names. The cell common to columns, *First Class* and *Survived* provide the count of first class passengers who survived, 203. Similarly, the cell common to columns *Crew* and *Died* provide the count of crew members who died on the ship, 673.

Now to convert all of these values to probabilities, we divide each of the numeric cells by the total count which in this table is 2201. The resulting probability table based on this data is shown as Table 3.

	Passenger Type				
	First Class	Second Class	Third Class	Crew	Total
Survived	0.092	0.054	0.081	0.096	0.323
Died	0.051	0.076	0.240	0.306	0.677
Total	0.148	0.130	0.321	0.402	1.0

Table 3: A probability table relevant to data about the ship *Titanic*

In a probability table, we define the probabilities in the cells near the margins as *marginal probabilities*. Note that the marginal probabilities (along an axis) should add up to 1.0, (e.g., $P(survived) + P(died) = 0.323 + 0.677 = 1.0$). We can verify the same if we sum up the marginal probabilities along the columns. The probability in the cell at the bottom-right corner is the probability of the sample space and therefore it is always equal to 1.0.

The other remaining cells all contain what we call *joint probabilities*. These contain the probabilities of the intersection of the events stated in the row and columns, for instance the top-left cell contains the probability that a first class passenger survived ($P(\textit{First Class} \cap \textit{Survived}) = 0.092 = 9.2\%$).

When dealing with problems that consist of multiple events, preparing such a probability table will be very useful.

4 Bayes' Rule

Given a probability $P(A|B)$, Bayes' rule can be used to calculate the "flipped" probability which is $P(B|A)$. We can derive the rule as follows:

$$P(A \cap B) = P(A) \times P(B|A)$$
$$\text{also, } P(A \cap B) = P(B) \times P(A|B)$$

Equating, we get

$$
\begin{aligned}
P(A) \times P(B|A) &= P(B) \times P(A|B) \\
P(A|B) &= P(A) \times P(B|A)/P(B)
\end{aligned} \tag{6}
$$

Similarly, we can also calculate

$$P(B|A) = P(B) \times P(A|B)/P(A) \tag{7}$$

So given one conditional probability, if we know the marginal probabilities we can easily calculate the other conditional probabilities.

From probability tables, we know

$$P(B) = P(A \cap B) + P(A^c \cap B)$$

From general multiplicative rule, get

$$P(B) = P(A) \times P(B|A) + P(A^c) \times P(B|A^c)$$

Substituting this value of $P(B)$ in (6), we get

$$P(A|B) = P(B|A) \times P(A)/P(B|A) \times P(A) + P(B|A^c) \times P(A^c)$$

This is the famous Bayes' rule for **two** events. So by knowing both $P(B|A)$ and $P(B|A^c)$ we can "flip" A and calculate $P(A|B)$

We can also scale up the Bayes' rule to more than two events. Let $A_1, A_2, \ldots, A_m$ be m mutually exclusive (disjoint) and totally exhaustive events. We can apply Bayes' rule as:

$$P(A_1|B) = \frac{P(B|A_1) \times P(A_1)}{P(B|A_1) \times P(A_1) + \cdots + P(B|A_m) \times P(B|A_m)}$$