This lecture will focus on Bayes Theorem, but first we must cover some concepts

1 Preliminaries

- First, consider the following situation:
  Let $A, B_1, B_2, ..., B_n$ be events of the sample space S such that
  1. $S = B_1 \cup B_2 \cup ... \cup B_n$
  2. $B_i \cap B_j = \varnothing \forall i, j,$ where $i \neq j$

  (i.e. $B_1, ..., B_n$ form a partition)

- Then, we see that

$$
\begin{aligned}
A &= A \cap S \\
&= A \cap (B_1 \cup B_2 \cup ... \cup B_n) \\
&= (A \cap B_1) \cup (A \cap B_2) \cup ... \cup (A \cap B_n) \leftarrow \text{By assumption} \\
\Rightarrow P(A) &= P((A \cap B_1) \cup (A \cap B_2) \cup ... \cup (A \cap B_n)) \\
\text{Note, By assumption we have that} & \\
B_i \cap B_j &= \varnothing \forall i, j, \text{ where } i \neq j \\
\Rightarrow (A \cap B_i) \cap (A \cap B_j) &= \varnothing \forall i, j, \text{ where } i \neq j \\
\text{Thus } A \cap B_i \text{ and } A \cap B_j \text{ are disjoint} & \quad \forall i, j, \text{ where } i \neq j \\
\text{Therefore} & \\
P(A) &= P((A \cap B_1) \cup (A \cap B_2) \cup ... \cup (A \cap B_n)) \\
&= P(A \cap B_1) + P(A \cap B_2) + ... + P(A \cap B_n)) \\
& \qquad \uparrow \text{By the fourth rule of probability} \\
&= \sum_{i=1}^{n} P(A \cap B_i)
\end{aligned}
$$

- Next, consider for the same sets as above, we see that $\forall i$

$$
\begin{aligned}
P(A \cap B_i) &= \frac{P(A \cap B_i)}{P(B_i)} P(B_i) \\
&= P(A|B_i) P(B_i)
\end{aligned}
$$

2 Bayes Rules

- Given set $A, B_1, ..., B_n$ as defined above, Bayes Rules is as follows:

$$P(B_i|A) = \frac{P(A|B_i)P(B_i)}{\sum_{j=1} P(A|B_j)P(B_j)}$$

- *Proof.*

$$
\begin{aligned}
P(B_i|A) &= \frac{P(B_i \cap A)}{P(A)} \\
&= \frac{P(B_i \cap A)}{\sum_{j=1}^{n} P(A \cap B_j)} \leftarrow \text{First result from above} \\
&= \frac{P(B_i \cap A)}{\sum_{j=1}^{n} P(A|B_j)P(B_j)} \leftarrow \text{Second result from above} \\
&= \frac{P(A|B_i)P(B_i)}{\sum_{j=1}^{n} P(A|B_j)P(B_j)} \leftarrow \text{From the defininton of conditional probability}
\end{aligned}
$$

□

- Bayes rule can be very useful determining the probability of an event in the past given the occurence of an event in the past

- To do this, suppose events $B_1, ...B_n$ represent possible events that happened in the past, and event A is an event that just happened. Then Bayes Rule can tell us the probability of one of those preceeding events, given that you know a later outcome

3 Example

Suppose that you have two cars, a large one and a small one. 3/4 of the time you drive the small car to work, and 1/4 of the time you drive the large one. When you drive the large car to work you have trouble getting a parking spot and thus are late to work 40% of the time. When you take the smaller car you are late only 10% of the time.
Given that you are on time on a particular morning, what is the probability that you drove the large car?
Solution:
So, what we want to know is $P(\text{Large Car}|\text{on time})$. In this case A is the event that you were on time, and B_1, B_2 are the events that you drove the large car and the event that you drove the small car. So, by Bayes rule:

$$
\begin{aligned}
P(\text{Large Car}|\text{on time}) &= \frac{P(\text{on time}|\text{Large Car})P(\text{Large Car})}{P(\text{on time}|\text{Large Car})P(\text{Large Car}) + P(\text{on time}|\text{Small Car})P(\text{Small Car})} \\
&= \frac{(.6)(1/4)}{(.6)(1/4) + (.9)(3/4)} \\
&\approx .182
\end{aligned}
$$

4 Exercise

Suppose that you are worried that you have cholera and decide to go get tested. In meeting with the doctor you are told that the test has a 5% False positive rate and a 10% false negative rate. Additionally, you are told that only 1% of the US population has cholera at a given time (Actually its much lower [almost no cases], but lets keep the numbers as nice as we can).

Suppose the test comes back positive, what is the probability that you have cholera?

Solution:

In this problem A is the even that the test came back positive and B_1 and B_2 are the events that you have cholera and don't have cholera, respectively. Then applying Bayes rule we see that

$$\begin{aligned} P(B_1|A) &= \frac{P(A|B_1)P(B_1)}{\sum_{i=1}^{2} P(A|B_i)P(B_i)} \\ &= \frac{P(\text{test positive}|\text{cholera})P(\text{cholera})}{P(\text{test positive}|\text{cholera})P(\text{cholera}) + P(\text{test positive}|\text{no cholera})P(\text{no cholera})} \end{aligned}$$

We have all of these values: $P(\text{cholera}) = 1/100$, $P(\text{no cholera}) = 1 - P(\text{cholera}) = 99/100$ (by the compliment rule), $P(\text{test positive}|\text{cholera}) = 9/10$ (false negative rate of 10%), and $P(\text{test positive}|\text{no cholera}) = 1/20$ (false positive rate of 5%). So, we can plug all of these values to get

$$\begin{aligned} P(\text{cholera}|\text{test positive}) &= \frac{P(\text{test positive}|\text{cholera})P(\text{cholera})}{P(\text{test positive}|\text{cholera})P(\text{cholera}) + P(\text{test positive}|\text{no cholera})P(\text{no cholera})} \\ &= \frac{(9/10)(1/100)}{(9/10)(1/100) + (1/20)(99/100)} \\ &= \frac{18/20}{18/20 + 99/20} \\ &= \frac{18}{117} \approx .15 \end{aligned}$$