4.1. Conditional Probability

The conditional probability of event A_1 given event A_2 is defined as follows

$$P(A_1 \mid A_2) = \frac{P(A_1 \cap A_2)}{P(A_2)}$$

Mathematically this formula amounts to making A_2 the new reference set, i.e., the set A_2 is now given probability 1 since

$$P(A_2 \mid A_2) = \frac{P(A_2 \cap A_2)}{P(A_2)} = 1$$

Example : Let a pair of fair dice be tossed. If the sum is 6, find the probability that one of the dice is a 2. In other words, if

$$E = \{\text{sum is } 6\} = \{(1,5), (2,4), (3,3), (4,2), (5,1)\}$$

and

$$A = \{\text{a 2 appears on at least one die}\}$$

find $P(A \mid E)$.

Now E consists of five elements and two of them, $(2,4)$ and $(4,2)$, belong to A: $A \cap E = \{(2,4), (4,2)\}$. Then $P(A \mid E) = \frac{2}{5}$.

On the other hand, since A consists of eleven elements,

$$A = \{(2,1), (2,2), (2,3), (2,4), (2,5), (2,6), (1,2), (3,2), (4,2), (5,2), (6,2)\}$$

and S consists of 36 elements, $P(A) = \frac{11}{36}$.

4.2. Independence of 2 Events

The notion of independence is crucial. Intuitively two events A_1 and A_2 are independent if knowing that A_2 has happened does not change the probability of A_1. In other words

$$P(A_1 \mid A_2) = P(A_1)$$

More generally we say that the events A and A_2 are independent if and only if

$$P(A_1 \cap A_2) = P(A_1).P(A_2)$$

4.3. Independence of n Events

We say that the events $A_1, \ldots, A_n$ are independent if and only if the following conditions are met:

1- All pairs of events with different indexes are independent, i.e.,

$$P(A_i \cap A_j) = P(A_i)P(A_j)$$

for all $i, j \in \{1, 2, \ldots, n\}$ such that $i = j$.

2- For all triplets of events with different indexes

$$P(A_i \cap A_j \cap A_k) = P(A_i)P(A_j)P(A_k)$$

for all $i, j, k \in \{1, \ldots, n\}$ such that $i = j = k$.

3- Same idea for combinations of 3 sets, 4 sets, . . .

4- For the n-tuple of events with different indexes

$$P(A_1 \cap A_2 \cap \cdots \cap A_n) = P(A_1)P(A_2) \cdots P(A_n)$$

Example : Let a fair coin be tossed three times; we obtain the equiprobable space

$$S = \{\text{HHH, HHT, HTH, HTT, THH, THT, TTH, TTT}\}$$

Consider the events

$$A = \{\text{first toss is heads}\}, \qquad B = \{\text{second toss is heads}\}$$

$$C = \{\text{exactly two heads are tossed in a row}\}$$

Clearly A and B are independent events; this fact is verified below. On the other hand, the relationship between A and C or B and C is not obvious. We claim that A and C are independent, but that B and C are dependent. We have

$$P(A) = P(\{\text{HHH, HHT, HTH, HTT}\}) = \frac{4}{8} = \frac{1}{2}$$

$$P(B) = P(\{\text{HHH, HHT, THH, THT}\}) = \frac{4}{8} = \frac{1}{2}$$

$$P(C) = P(\{\text{HHT, THH}\}) = \frac{2}{8} = \frac{1}{4}$$

Then

$$P(A \cap B) = P(\{\text{HHH, HHT}\}) = \frac{1}{4}, \quad P(A \cap C) = P(\{\text{HHT}\}) = \frac{1}{8},$$

$$P(B \cap C) = P(\{\text{HHT, THH}\}) = \frac{1}{4}$$

4.4. The Chain Rule of Probability

Let $\{A_1, A_2, . . . , A_n\}$ be a collection of events. The chain rule of probability tells us a useful way to compute the joint probability of the entire collection

$P(A_1 \cap A_2 \cap \cdots \cap A_n) =$

$$P(A_1)P(A_2|A_1)P(A_3|A_1 \cap A_2) \ldots P(A_n|A_1 \cap \ldots\ldots A_{n-1}) =$$

$$P(A_1)\frac{P(A_2 \cap A_1)}{P(A_1)}\frac{P(A_3 \cap A_2 \cap A_1)}{P(A_1 \cap A_2)} \ldots \frac{P(A_1 \cap \ldots .\cap A_n)}{P(A_1 \cap \ldots .\cap A_{n-1})} = P(A_1 \cap \ldots .\cap A_n)$$

Example: A car company has 3 factories. 10% of the cars are produced in factory 1, 50% in factory 2 and the rest in factory 3. One out of 20 cars produced by the first factory are defective. 99% of the defective cars produced by the first factory are returned back to the manufacturer. What is the probability that a car produced by this company is manufactured in the first factory, is defective and is not returned back to the manufacturer.

Let A_1 represent the set of cars produced by factory 1, A_2 the set of defective cars and A3 the set of cars not returned. We know

$$P(A_1) = 0.1$$

$$P(A_2 \mid A_1) = 1/20$$

$$P(A_3 \mid A_1 \cap A_2) = 1 - 99/100$$

Thus, using the chain rule of probability

$$P(A \cap A_2 \cap A_3) = P(A_1)\, P(A_2 \mid A1)P(A_3 \mid A_1 \cap A_2) =$$

$$(0.1)\ \ (0.05)\ (0.01) = 0.00005$$

Example : We are given three boxes as follows:

Box I has 10 light bulbs of which 4 are defective.

Box II has 6 light bulbs of which 1 is defective.

Box III has 8 light bulbs of which 3 are defective.

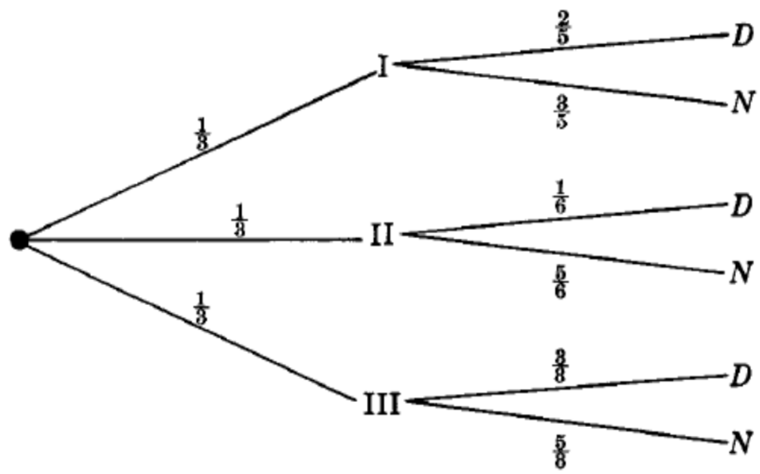
We select a box at random and then draw a bulb at random. What is the probability p that the bulb is defective?

Here we perform a sequence of two experiments:

(i) select one of the three boxes;

(ii) select a bulb which is either defective (D) or nondefective (N).

The following tree diagram describes this process and gives the probability of each branch of the tree:



The probability that any particular path of the tree occurs is, by the multiplication theorem, the product of the probabilities of each branch of the path, e.g., the probability of selecting box I and then a defective bulb is $\frac{1}{3} \cdot \frac{2}{5} = \frac{2}{15}$.

Now since there are three mutually exclusive paths which lead to a defective bulb, the sum of the probabilities of these paths is the required probability:

$$p = \frac{1}{3} \cdot \frac{2}{5} + \frac{1}{3} \cdot \frac{1}{6} + \frac{1}{3} \cdot \frac{3}{8} = \frac{113}{360}$$

4.5. Bayes' Theorem

This theorem, which is attributed to Bayes (1744-1809), tells us how to revise probability of events in light of new data. It is important to point out that this theorem is consistent with probability theory and it is accepted by frequentists and Bayesian probabilists. There is disagreement however regarding whether the theorem should be applied to subjective notions of probabilities (the Bayesian approach) or whether it should only be applied to frequentist notions (the frequentist approach).

$$P(H_i \mid D) = \frac{P(D \mid H_i)P(H_i)}{P(D \mid H_1)P(H_1) + P(D \mid H_2)P(H_2) + \cdots}$$

where

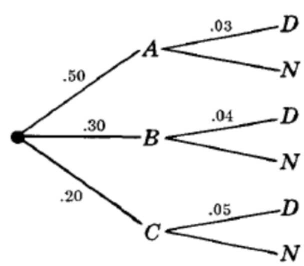
- $P(H_i)$ is known as the prior probability of the hypothesis H_i. It evaluates the chances of a hypothesis prior to the collection of data.
- $P(H_i \mid D)$ is known as the posterior probability of the hypothesis H_i given the data.
- $P(D \mid H_1)$, $P(D \mid H_2)$, . . . are known as the likelihoods.

Example : Three machines A, B and C produce respectively 50%, 30% and 20% of the total number of items of a factory. The percentages of defective output of these machines are 3%, 4% and 5%. If an item is selected at random, find the probability that the item is defective.

Let X be the event that an item is defective. Then by (*1*) above,

$$\begin{aligned} P(X) &= P(A)\,P(X \mid A) + P(B)\,P(X \mid B) \\ &\qquad + P(C)\,P(X \mid C) \\ &= (.50)(.03) + (.30)(.04) + (.20)(.05) \\ &= .037 \end{aligned}$$

Observe that we can also consider this problem as a stochastic process having the adjoining tree diagram.



Example : Consider the factory in the preceding example. Suppose an item is selected at random and is found to be defective. Find the probability that the item was produced by machine A; that is, find $P(A \mid X)$.

By Bayes' theorem,

$$P(A \mid X) = \frac{P(A)\,P(X \mid A)}{P(A)\,P(X \mid A) + P(B)\,P(X \mid B) + P(C)\,P(X \mid C)}$$

$$= \frac{(.50)(.03)}{(.50)(.03) + (.30)(.04) + (.20)(.05)} = \frac{15}{37}$$

In other words, we divide the probability of the required path by the probability of the reduced sample space, i.e. those paths which lead to a defective item.

<Best Regards>