Chapter I: Probability

Probability is the mathematical quantification of uncertainty.

The probability theorem consists of three major things: ***Sample Space, Outcomes**, and **Events**.*

Sample space is the set of all possible outcomes of a standard experiment. Each point in the set are the **outcomes** of the experiment. An **event** is a subset of the sample space that can contain one or more outcomes.

Example: Tossing a coin twice will create a sample space of $\Omega = \{HH, HT, TH, TT\}$ these possible outcomes. Here $\{HT\}$ and $\{TT\}$ are both separate outcomes, and also are separate subsets of Ω. The probability of the first toss is heads is called an event (A), which is also a subset of Ω; $A = \{HH, HT\}$.

The complement of an event is denoted as A^C, which means it contains all outcomes in the sample space that are not in event A.

A **sure event** is Ω itself (probability=1) and the **null event** ϕ is the empty set (probability=0). Any event is a subset of Ω.

For a finite sample space, $\Omega = \{\omega_1, \omega_2, \omega_3..., \omega_n\}$, if all outcomes are equally likely, then the probability of event A can be denoted as: $\mathbb{P}(A) = \frac{|A|}{|\Omega|}$, which is called the **uniform probability distribution.**

Independent Events

If tossing a fair coin twice has a sample space of {HH, HT, TH, TT}, then the probability of two heads could be determined as follows:

The first toss has half ($1/2$) the probability of heads. Then the second toss has the same half ($1/2$) probability of heads. We consider the two tosses separate, or more specifically, independent. Therefore, the probability of two heads will be multiplied, and the results would be $(1/2 \times 1/2) = 1/4$.

Conditional Probability *(The base of Bayes' Theorem)*

The probability of an event occurring, given that another event is already known to have occurred, is called the conditional probability. Let assume that event B has already occurred, and the probability of event B is greater than zero, $\mathbb{P}(B) > 0$. Then the conditional probability of event A given B is: $\mathbb{P}(A|B) = \frac{\mathbb{P}(A \cdot B)}{\mathbb{P}(B)}$.

Example: A medical test for a disease D has outcomes + (positive) and – (negative). The given probabilities are:

	D	D^C
+	0.009	0.099
–	0.001	0.891

Now, from the conditional probabilistic definition,

Conditional probability of + given D is: $\mathbb{P}(+|D) = \frac{\mathbb{P}(+\cap D)}{\mathbb{P}(D)} = \frac{0.009}{0.009+0.001} = 0.9$

and,

Conditional probability of – given D^C is: $\mathbb{P}(-|D^C) = \frac{\mathbb{P}(-\cap D^C)}{\mathbb{P}(D^C)} = \frac{0.891}{0.891+0.099} = 0.9$

But here is a catch. What if you go for a test and get a positive (+) result, then what is the probability of you having the disease? You may say it's the same as the result is +ve given that you have the disease, which is 0.9. But the correct answer is:

Conditional probability of D given + is: $\mathbb{P}(D|+) = \frac{\mathbb{P}(+\cap D)}{\mathbb{P}(+)} = \frac{0.009}{0.009+0.099} = 0.08.$

A Lemma from Conditional Probability (A base for Bayes'):

If A and B are independent events then, $\mathbb{P}(A|B) = \mathbb{P}(A)$.

Also, for any pair of events A and B: $\mathbb{P}(AB) = \mathbb{P}(A|B)\,\mathbb{P}(B) = \mathbb{P}(B|A)\,\mathbb{P}(A)$.

Example: Draw two cards from a deck, without replacing it. Let A be the event that the first draw is the Ace of Clubs (♣) and let B the event that the second draw is the Queen of Diamonds (◆). Then, $\mathbb{P}(AB) = \mathbb{P}(A)\,\mathbb{P}(B|A) = (1/52) \times (1/51)$.

Bayes' Theorem

Bayes' theorem is the basis of "expert systems" and "Bayes' nets" which are considered to be mimicking the decision-making ability of a human expert in a specific area. **Bayesian Networks** (Bayes' Nets) are a probabilistic graphical model that represents relationship between variables using a directed graph and do reasoning inferences based on available evidence.

The Law of Total Probability: Let $A_1,\ldots,A_k$ be a partition of Ω. Then for any event B,

$$\mathbb{P}(B) = \mathbb{P}(B|A_1)\mathbb{P}(A_1) + \cdots + \mathbb{P}(B|A_k)\mathbb{P}(A_k)$$

$$\mathbb{P}(B) = \sum_{i=1}^{k} \mathbb{P}(B|A_i)\mathbb{P}(A_i)$$

Here, we call:

$\mathbb{P}(A_i)$ the prior probability, or the *Priori* of A.

$\mathbb{P}(A_i|B)$ the posterior probability, or the *Posteriori* of A.

$$\mathbb{P}\left(A_i|B\right) = \frac{\mathbb{P}\left(A_i B\right)}{\mathbb{P}(B)} = \frac{\mathbb{P}\left(B|A_i\right)\mathbb{P}\left(A_i\right)}{\mathbb{P}(B)} = \frac{\mathbb{P}\left(B|A_i\right)\mathbb{P}\left(A_i\right)}{\sum_j \mathbb{P}\left(B|A_j\right)\mathbb{P}\left(A_j\right)}$$