**PRE-REQUISITE**

P(x) -> defines the probability of a random variable x

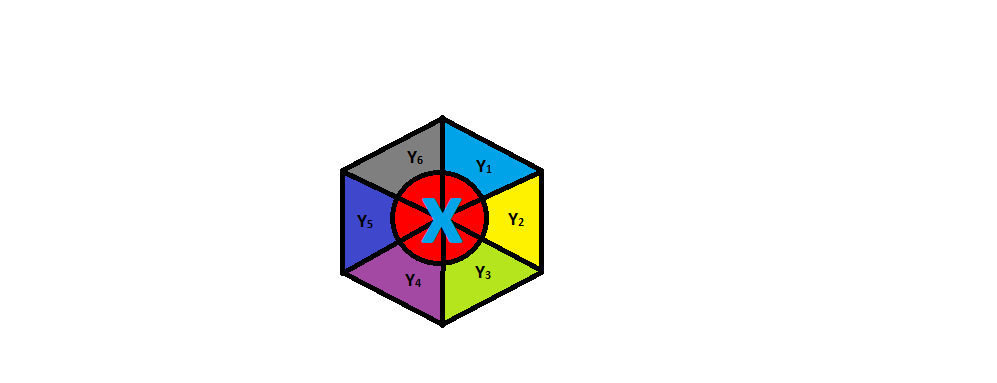
P(x|y) -> defines the probability of a random variable x provided y has already happened

P(x|y) = P(y|x) P(x) / P(x) ………Bayes theorem

**1** . ***THEOREM OF TOTAL PROBABILITY***

Let y1 ....yn  be a set of mutually exclusive events and events x is the union of N mutually exclusive events, then

P(x) = ∑ P(x|yi)P(y)



**2.EXPECTATION OF A VARIABLE X ……..E(x)**

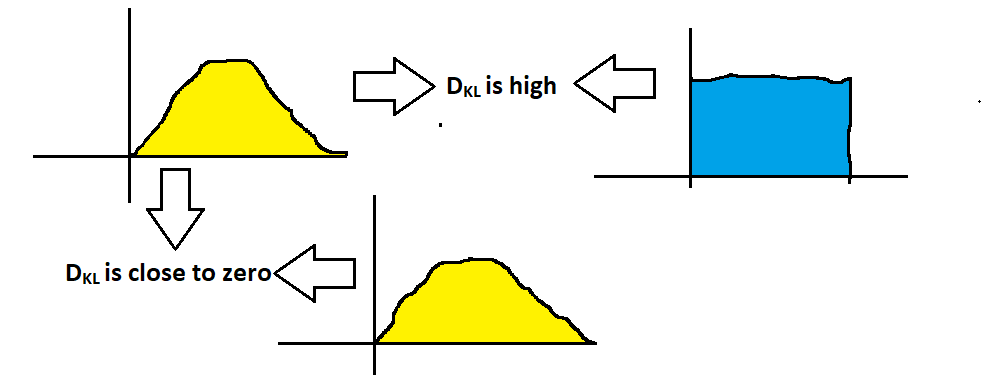
It is a weighted average of the possible values that x can take, each value being weighted according to the probability of that event defined as

E(x) = ∑ Xi P(Xi)

**3. KL DIVERGENCE**

Kullback –Leibler divergence is a measure of how one probability distribution P & Q, the KL divergence between them is defined as

DKL(P||Q) = ∑ P(x) log (P(x) /Q(x))

*PROPERTIES:*

* KL(P||Q) or KL(P||Q) >= 0
* KL(P||Q) != KL(Q||P) (not symmetric)