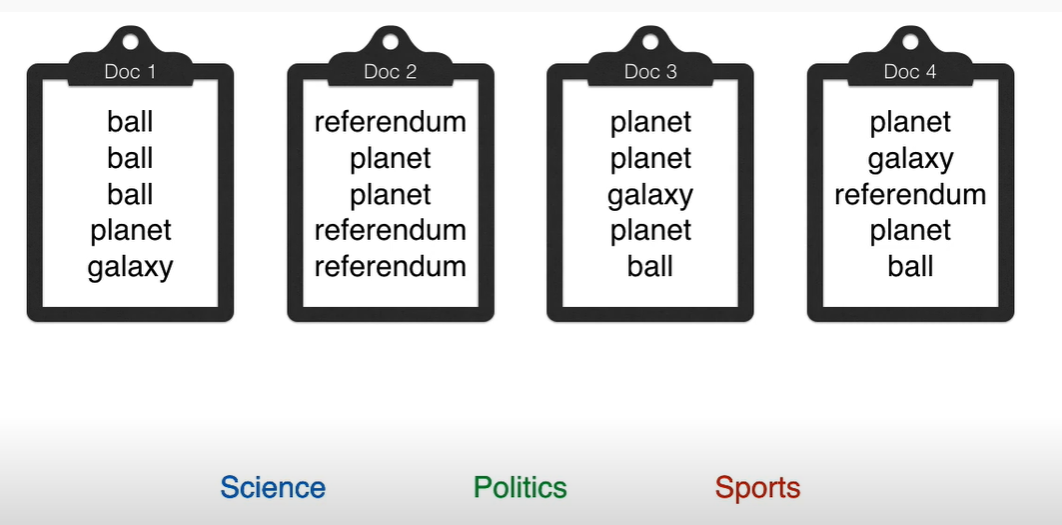
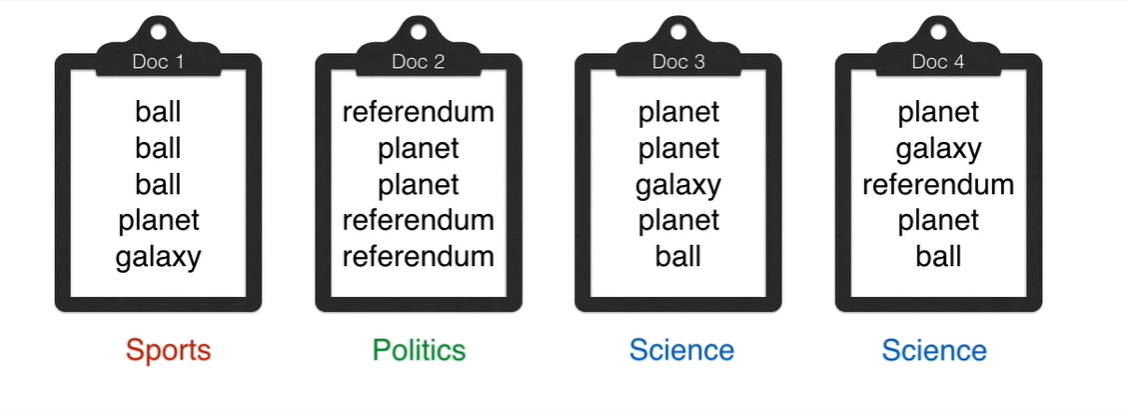
**Latent Dirichlet Allocation**

Latent Dirichlet Allocation is a Topic modelling methodology used in Machine learning and Natural Language Processing.

Let’s Assume that we have a set of documents. Each document is made up of statements formed using four words (ball, galaxy, planet and referendum). These documents cover three topics (Science, Sports and Politics)

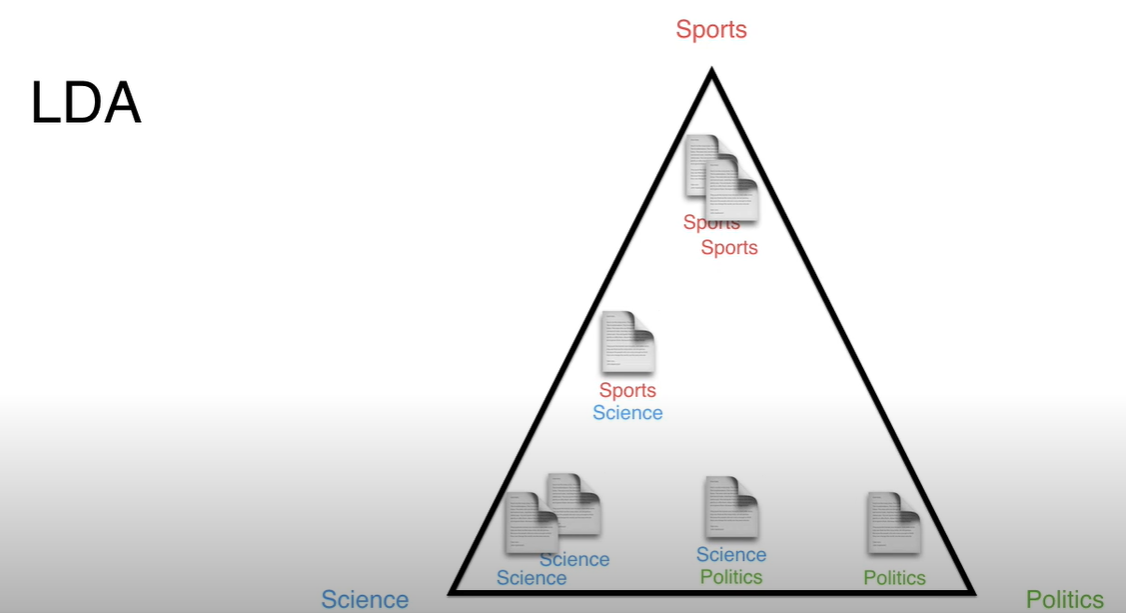


Now how does a human classify the documents. We find the majority of topic related words in each document. So, they can be classified like this.

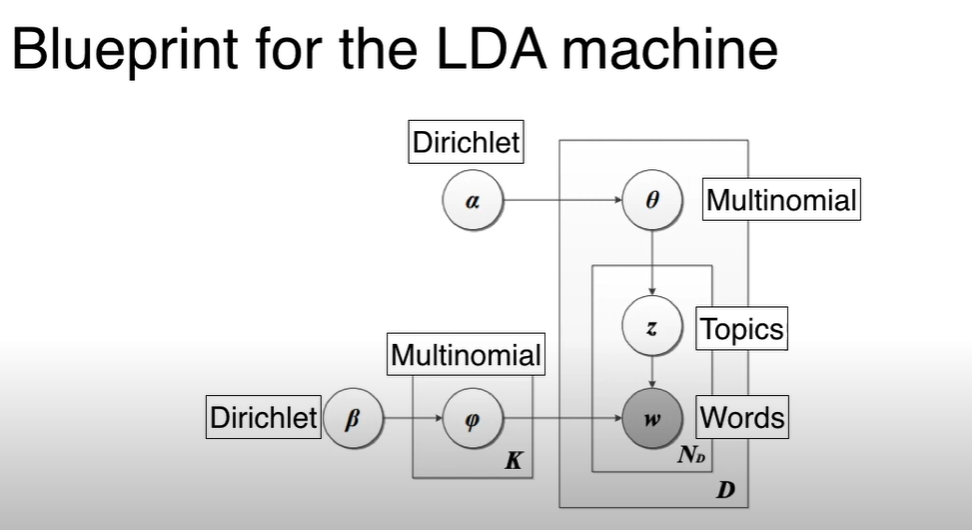


Now how will the computer do it?

LDA builds a triangle as shown below.



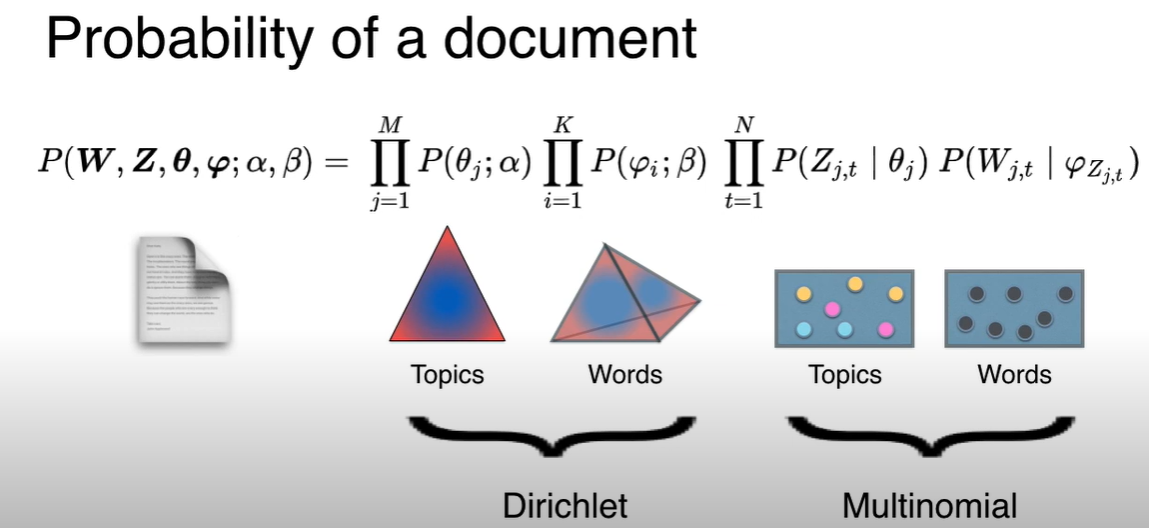
Here is the blueprint of a LDA Machine



As we can see, there are Dirichlet and Multinomial Distributions working here. Now lets simplify the blue print with the formula.

**Application of LDA Mathematically**

Finding the probability of the document being close to the required output is found out. Its like a score to determine the best model. We use the following calculation.

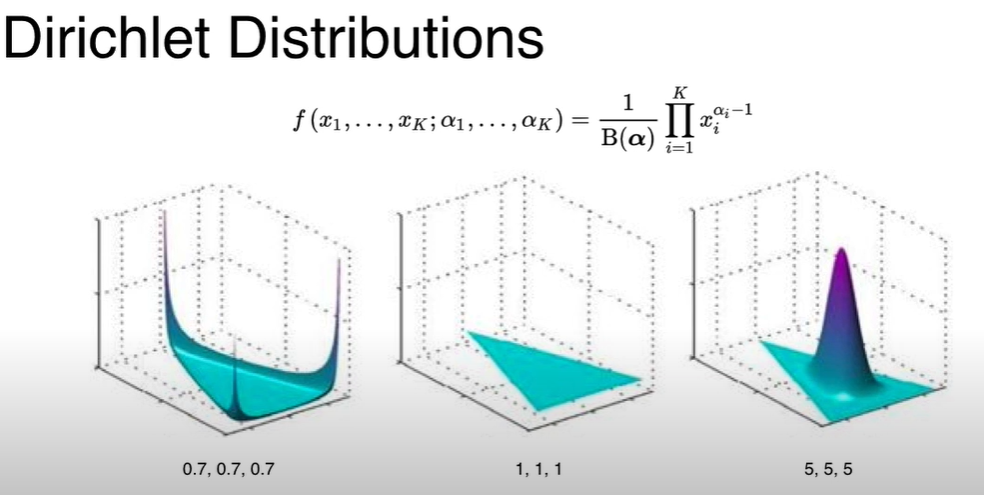


The first two calculations will give us the Dirichlet Distribution of Topics and words in the document respectively.

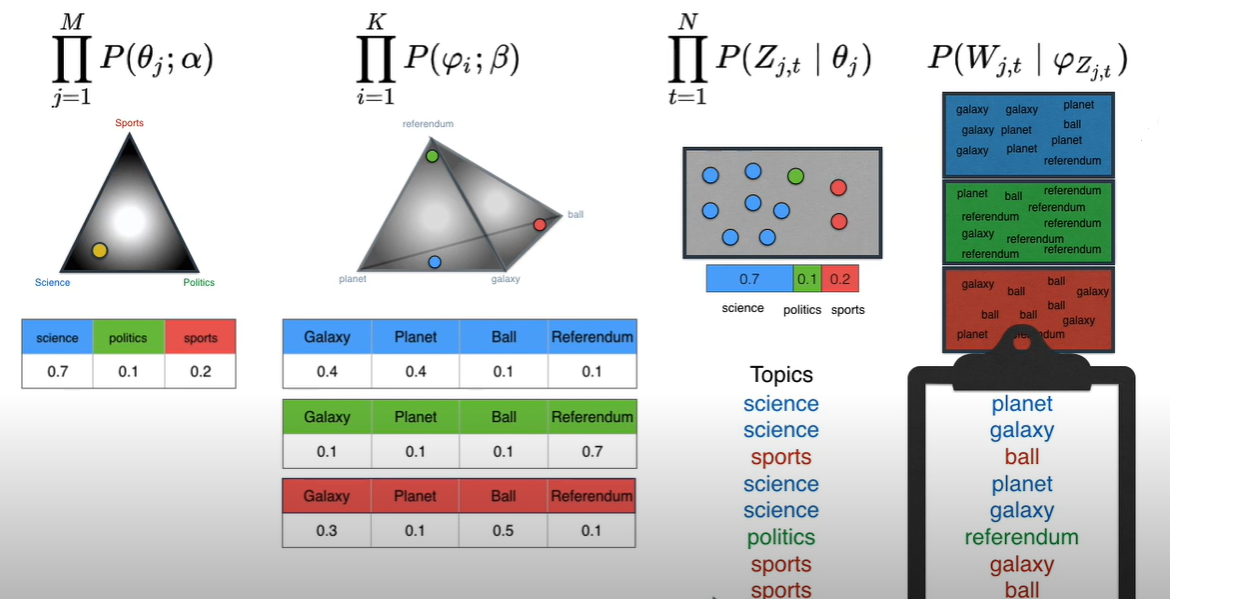
The last two generate the document by using multinomial distribution of the topics and words.

**Dirichlet Distributions**

Dirichlet distributions are non-continuous distributions. Here is the representation of Dirichlet distribution.



**Simple working of LDA**



**Now we can explain the blue print using the same formula.**

