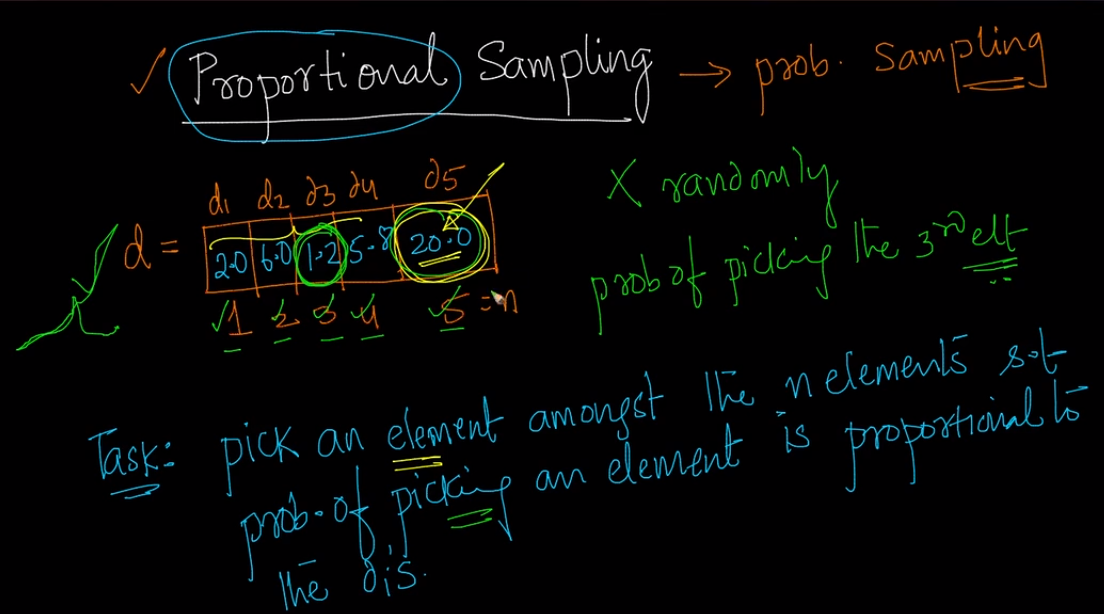
Proportional Sampling: It is used in many places or in many ml models.

It’s says that given ‘n’ element the probability of picking any element should be proportional to that value which is picking up.

Let’s take an example in below fig, we are given 5 elements (d), now probability of picking any random element from d is equal, but we want the probability to be proportional to di’s,

that means for given d the probability of picking up 5th element which is 20, should be highest or more as compare to other elements as value of 20 is comparatively larger than all values (that means while picking any element randomly, the 5th element should be picked more often).

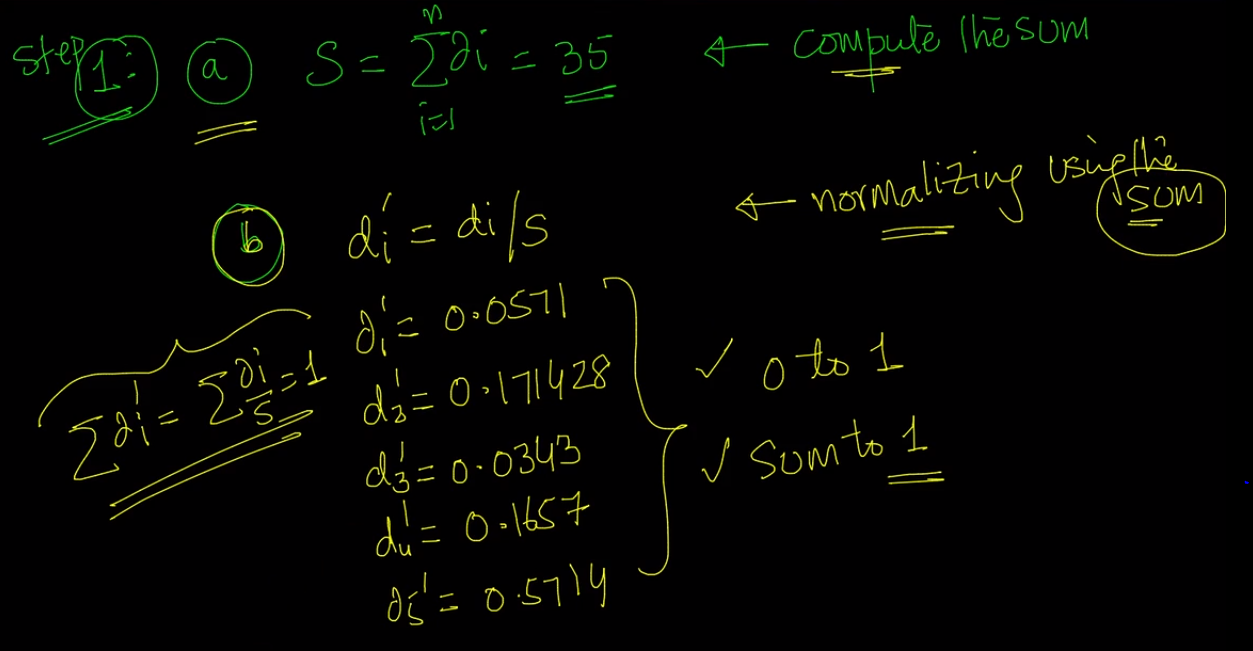
We can also say that probability of picking 3rd element which is 1.2 should be lowest, as 1.2 is very less as compare to other elements (that means while picking any element randomly, the 3rd element should be picked very less times as compare to other elements).



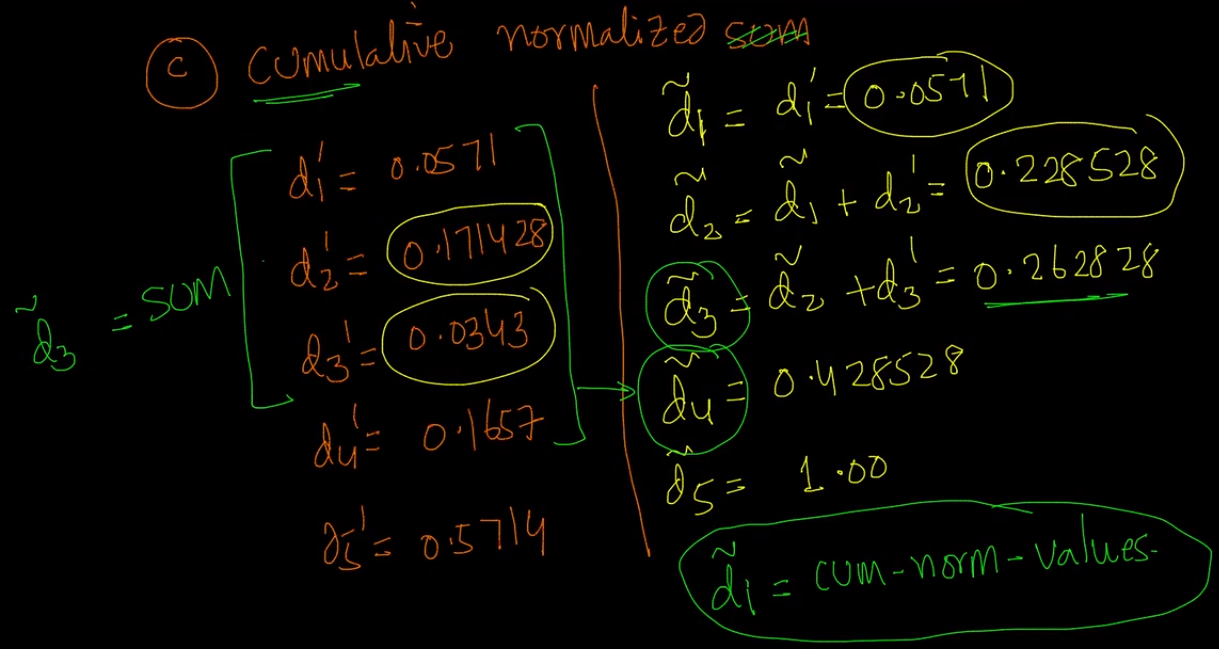
How we can achieve this ?

Step 1:

1. Take sum of all di’s, let’s say sum to be S.
2. Divide each di’s by sum ie di’ = di/S, in this step basically we are normalizing each value so that each value can come within 0 and 1, and if we take sum or all di’ will be equal to 1.

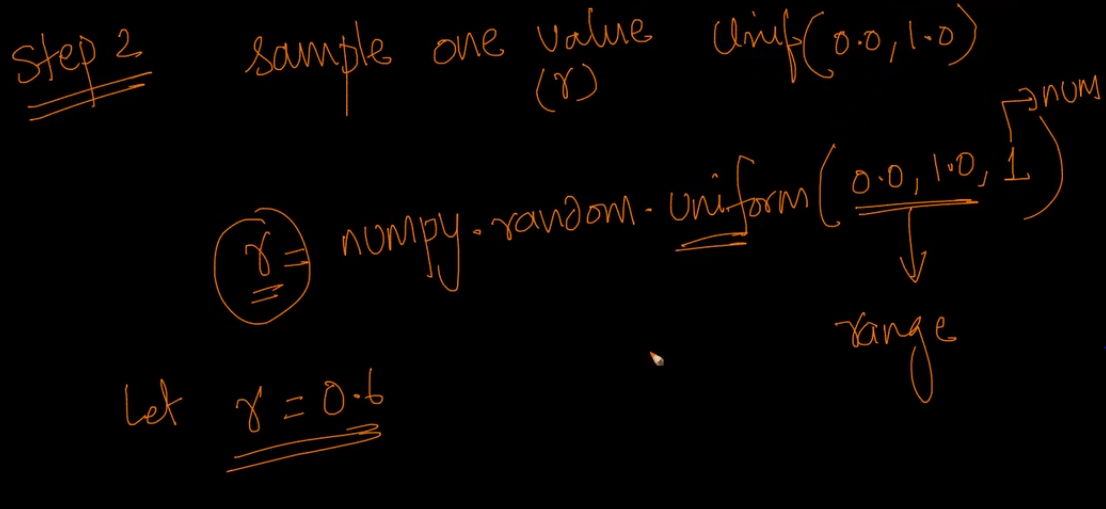


1. Take cumulative normalized: means summing up all the previous value upto that di



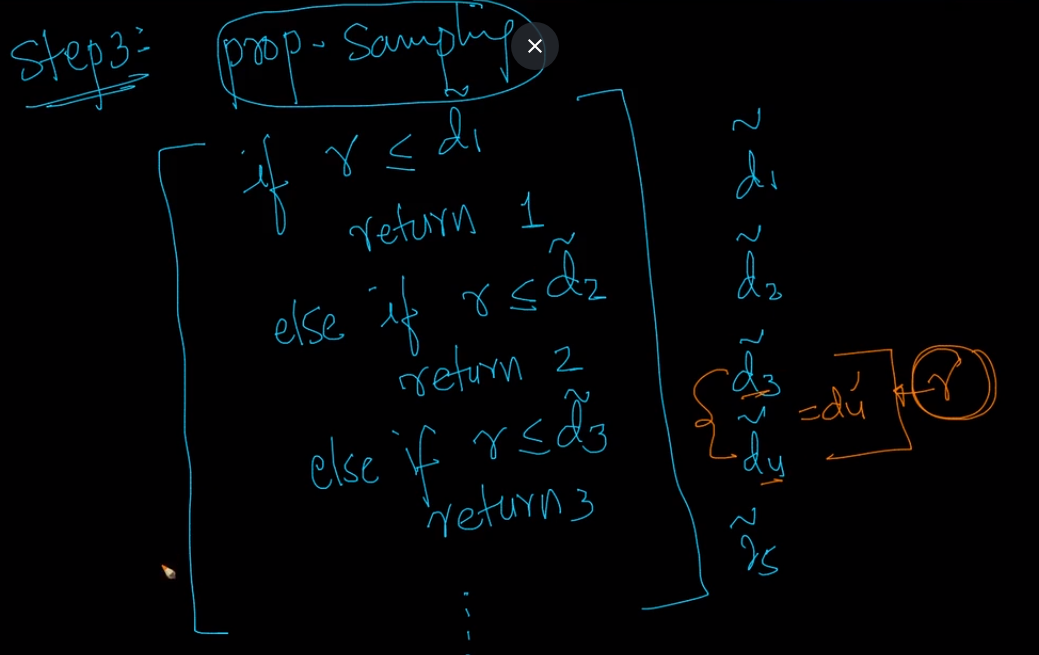
**Step 2:** Sample one value uniformly between 0.1 to 1.0, that means pick any 1 value uniformly within 0.0 to 1.0, we can get this using **uniform** function present in **random** package of **numpy.**

Let’s say we pick value r = 0.6



**Step 3:** In this step the proportional sampling will happen, as:

If **r** value find in step 3 is smaller or equal to d~i then pick ith  element.



Why this is so, basically we can say that the larger difference between current and previous element the more the chance of picking th current element.

Now how probability of picking element is proportional to value of that element.

Example:

Prob of picking up 4th element = probability of **r** lying between d~3 & d~4 = d4’

d4’ is proportional to d4.

