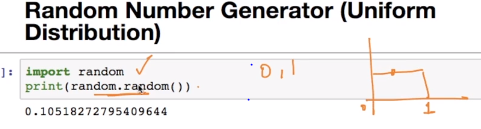
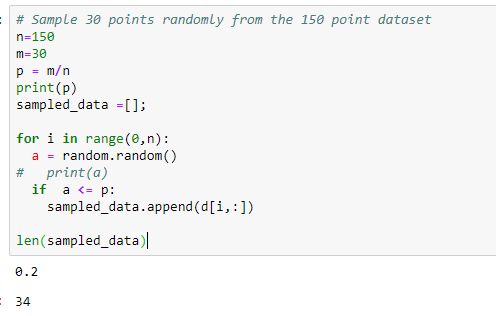
This chapter show how to randomly sample in uniform distributed way.

Any random generator will always generates the uniformly distributed numbers.



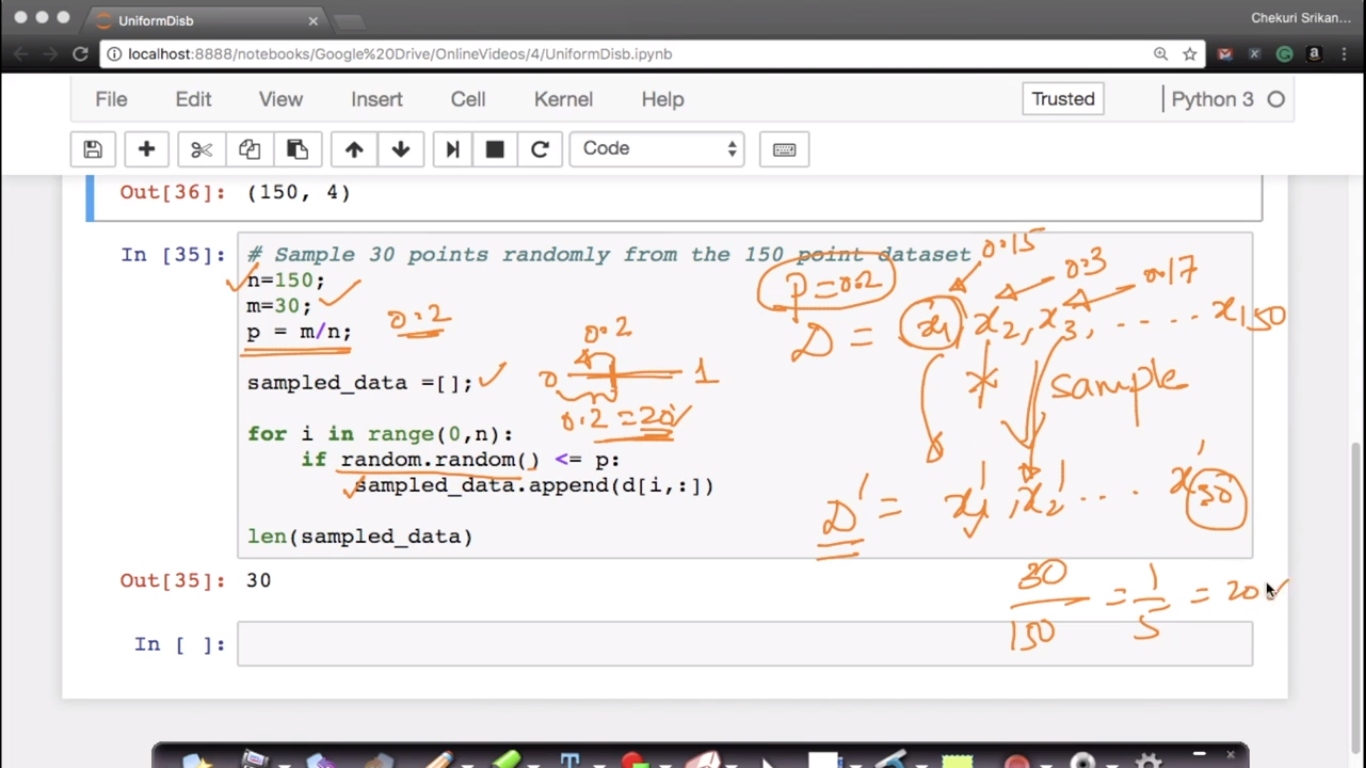
The idea is to pick 30 sample data from 150 point dataset where each data point we are getting is different from other, that is sampled data uniformly distributed.



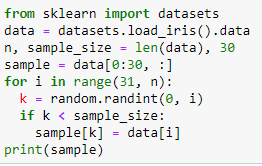
Here we need to take 30 out of 150, therefore probability for occurring any point in the sample is 0.2. Now we run loop from 1 to 150 and generates a random number, if random number is less than 0.2, then we’ll take the current iterating observation.

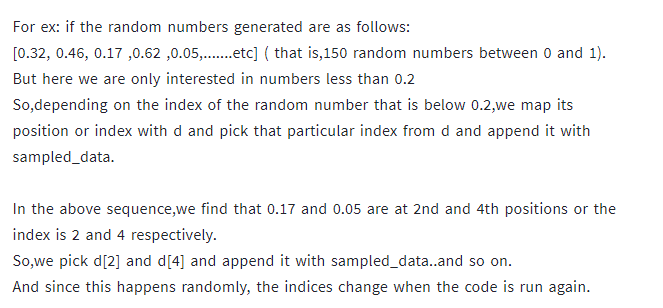
Note: the length of sampled\_data is not always 30, because in some case random generator might generate only 20 points that are less than 0.2 and it might possible it will generate 0 points less than 0.2, so in this case we should execute code until we get sample size less than 30.

* if D is 150 and D' is 150, then each point from D has probability of 1(150/150) to be present in D'.
* if D is 150 and D' is 75, only half of the points will make it to D', so for each point the probability will be 50-50 (.5) to make to D'. then each point from D has probability of 1(75/150) to be present in D'.
* if D is 150 and D' is 30, only few of the points will make it to D', so for each point the probability will be < (.5) to make to D'. then each point from D has probability of 1(30/150) to be present in D'.



**Example:** let’s take a example of iris data set where we want to take sample data from any of those 150 observations randomly but should not be repeated, that means it should be uniformly distributed.





So, Eventually we are getting 30 sample values randomly taken from any place in those 150 datasets.