Introduction

What is CLT?

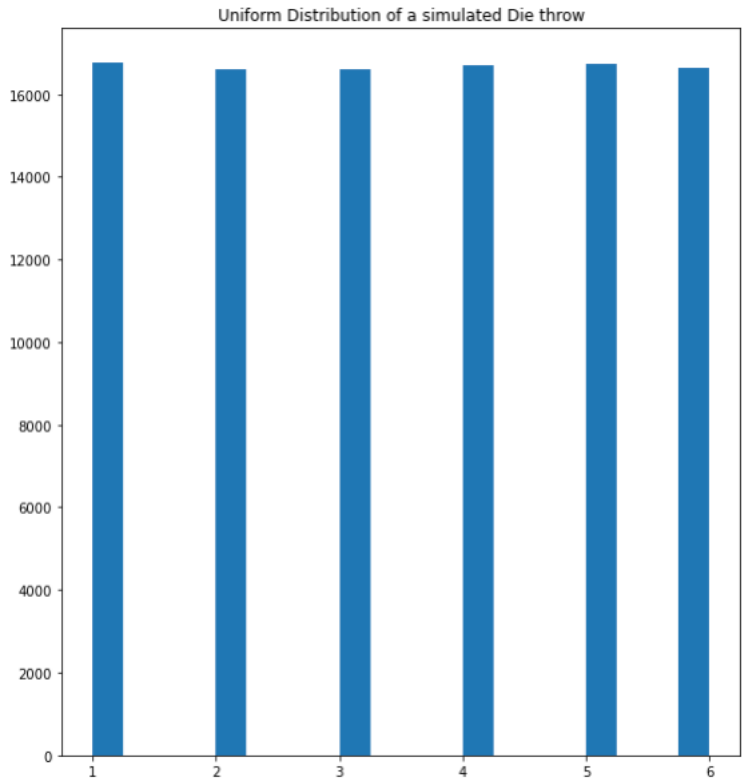
If we randomly take samples from a population, with sample size which is sufficiently big, the sample means will follow an approximately normal distribution with the mean of that distribution equal to the population mean.

Why CLT?

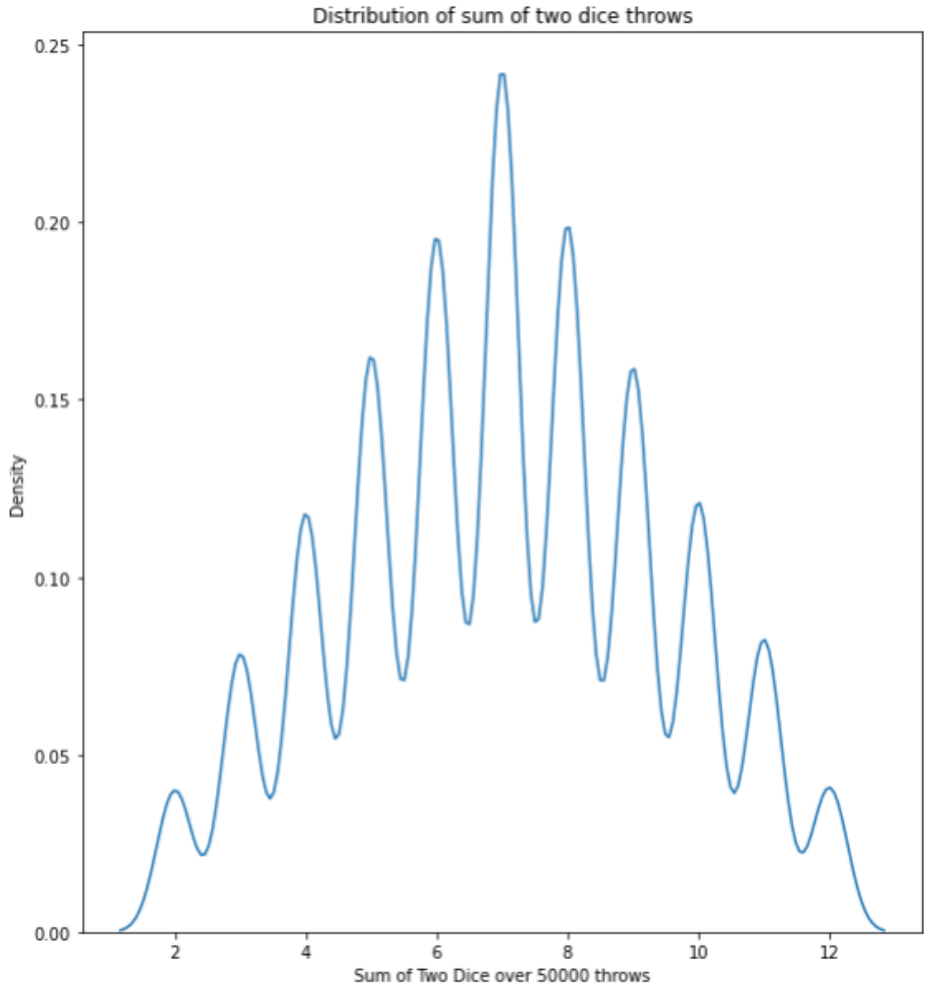
1. Used when population is too large to find mean directly.
2. The population doesn't have a clear distribution and we still want to know its mean.

Building Up the Population

1.Distribution obtained when one die is rolled.



2.Distribution of the sum of two dice: This will be taken as our population



The underlying distribution of the population is not normal, and it is a multimodal distribution.

Simulations for CLT

num_of_faces = 6

By default, this variable has been set to 6, but the simulations can be done for a hypothetical die with different number of faces also.

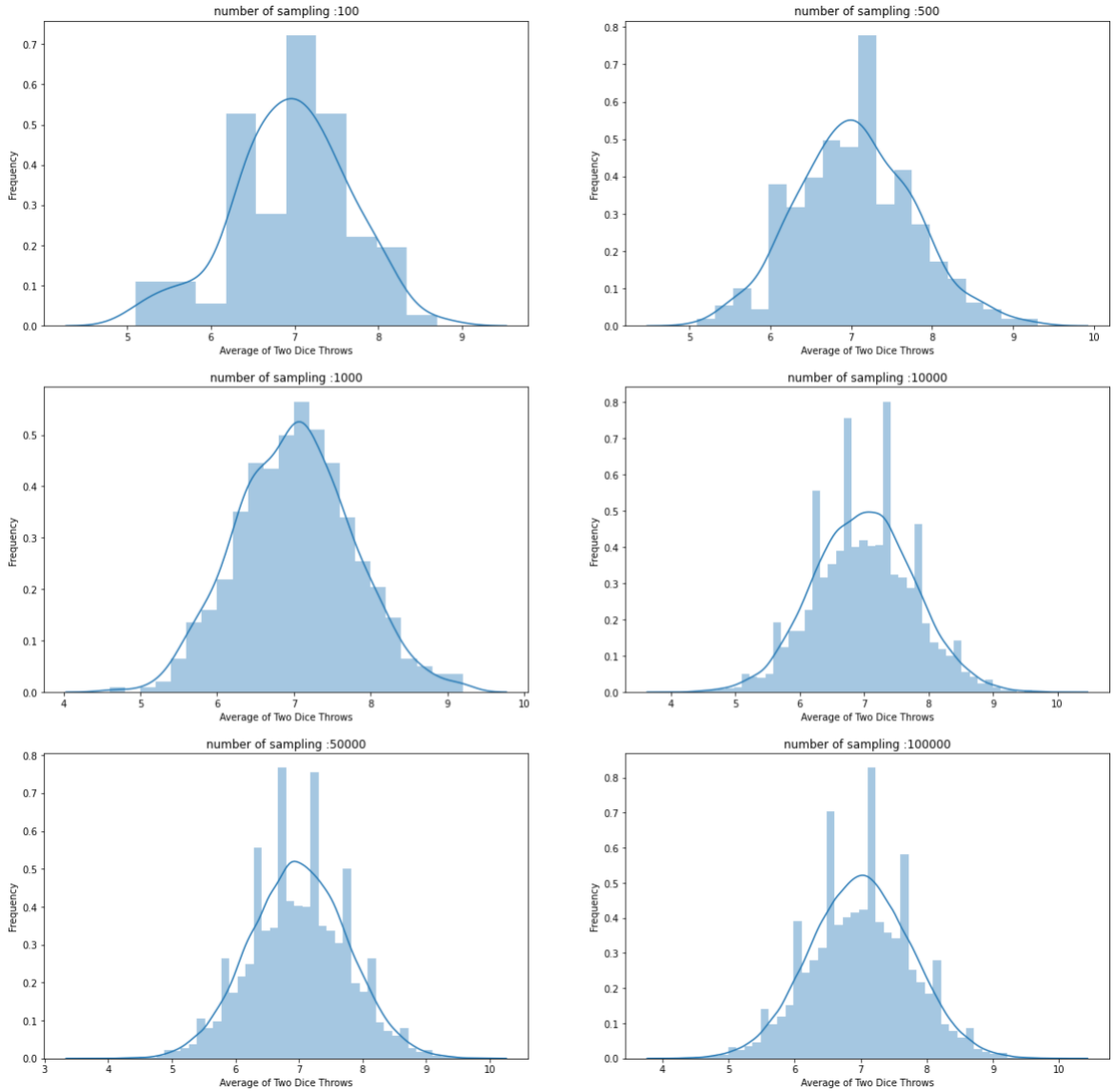
```
num_of_experiments = [
   100,
   500,
   1000,
   10000,
   50000,
   100000,
]
```

The experiment was simulated for different number of times (different number of samples).

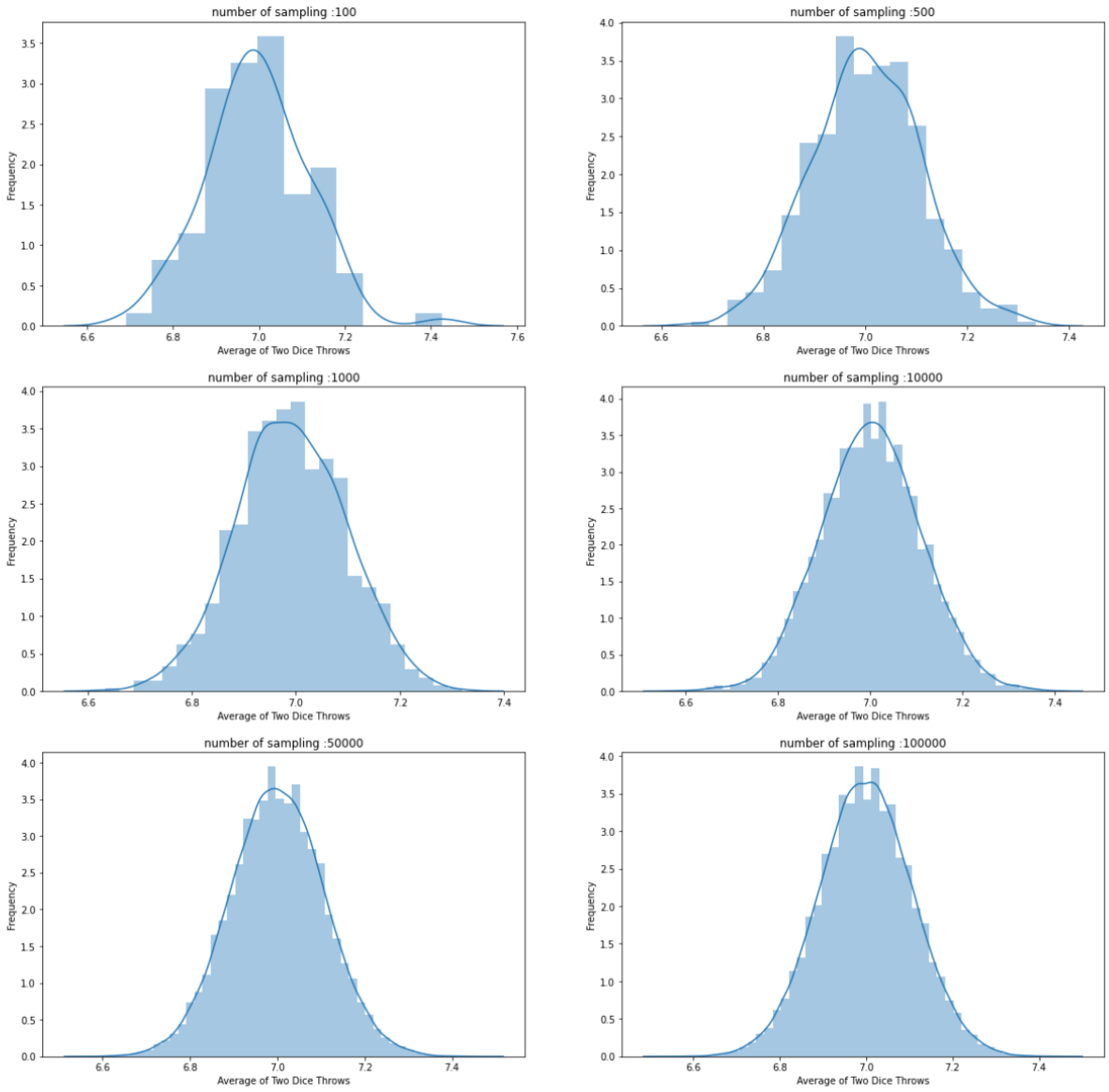
```
num_throws_per_experiment = 10
```

num_throws_per_experiment, is like sample size, and it has been set to a smaller number, to see what happens for a low sample size. (low meaning, n not greater than 30)

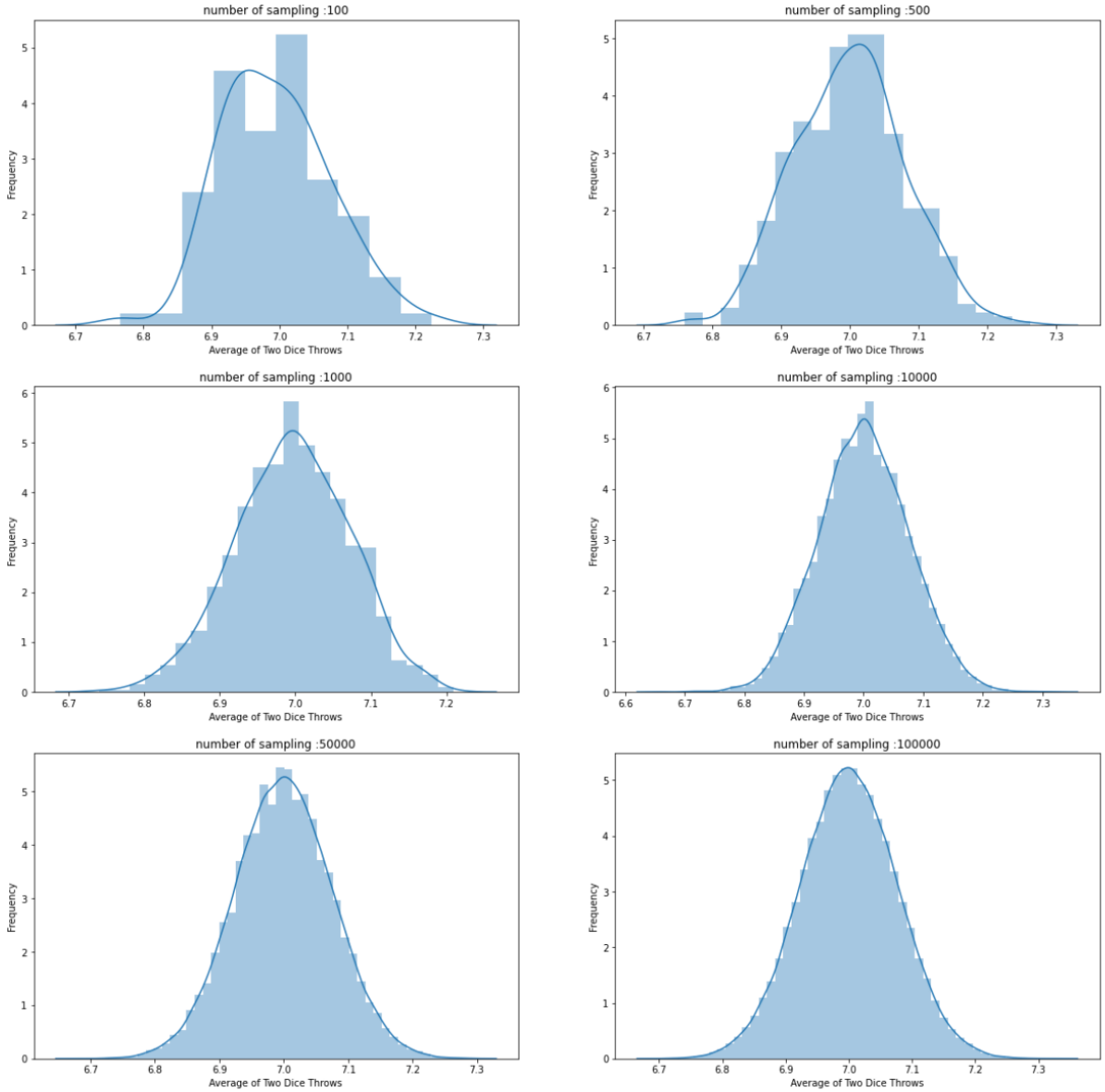
For a low sample size, the following can be seen in the graphs:
1.As the number of samples is lower (number of experiments) is low, the graph significantly deviates from normal distribution.
2. Increasing the number of samples does make it closer to normal distribution but it still shows some erratic behaviors like peaks and skewness.



The first set of 6 distributions is for num_throws_per_experiment =10



The second set of 6 distributions is for num_throws_per_experiment =500



The third set of 6 distributions is for num_throws_per_experiment =1000

For num_throws_per_experiment (sample size)=500 and 1000, a similar trend can be seen. As we increase the number of samples (number of experiments), the graph becomes closer to normal distribution, and for num_throws_per_experiment(sample size=500),the distribution gets very close to normal distribution as the number of samples(num_of_experiments) increase, however it is still not very smooth. However for num_throws_per_experiment = 1000, as we increase the number of samples the graph almost becomes like a perfect normal distribution.

Conclusion

1. As the number of experiments(number of samples) gets larger, the sample mean distribution gets closer to normal distribution.
2. As num_throws_per_experiments(sample size) gets bigger, the sample mean distribution gets closer to normal distribution.
3. The minimum sample size matters: only a sample size of at least 30 can ensure the mean follows the normal distribution.