Asymptotic Equipartition Property

Gantavya Bhatt

October 10, 2019

1 Problem Statement

- Generate a large number of samples independently from a Bernoulli distribution and store the sequence.
- Repeat the above procedure, and collect the sequences. Obtain the Histogram of the distribution of the sequences after collecting a large number of sequences.
- As you vary the value of p, observe any structure in the family of sequences you have generated.
- Choose a value of ϵ to define your own typical set. As you increase the value of n, observe the fraction of the sequences that lie in the typical set.

2 Introduction

This is one of the most important property in the Information theory. Before jumping to the results, we would like to state some expressions that led to the results.

2.1 Entropy

$$H = -\sum_{x \epsilon \chi} p(x) log(p(x))$$

For a Bernoulli distribution, this expression reduces to :-

$$H = -(plogp + (1-p)log(1-p))$$

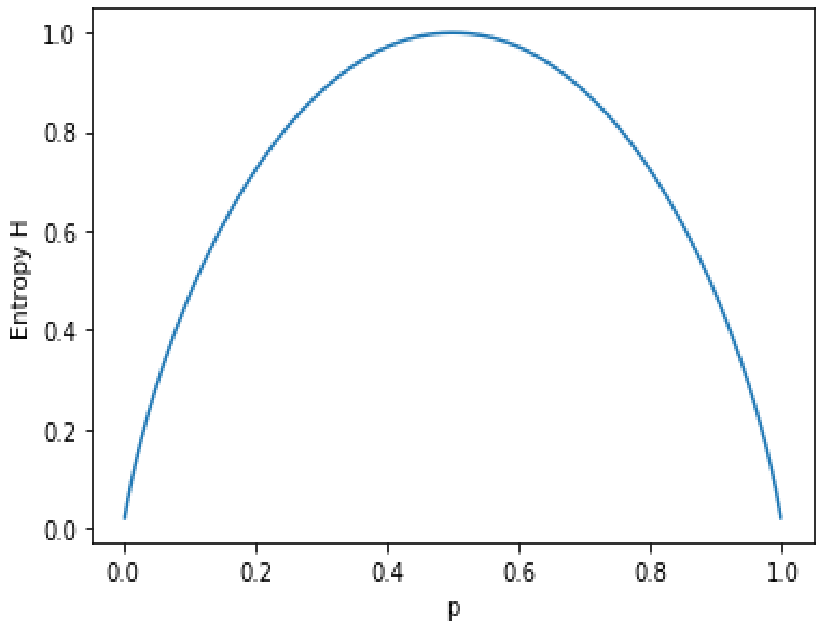

Figure 1: Entropy vs p value

2.2 Weak Law of Large Number

In probability, the sample mean converges to the true first order moment of the random variable. This can be seen from the limit perspective.

$$lim_{N \to \infty} \frac{\sum_{i=1}^{N} X_i}{N} = E(X)$$

Experimentally simulating this but for the different expression where X is replaced by log(p(X)) we get the following results.

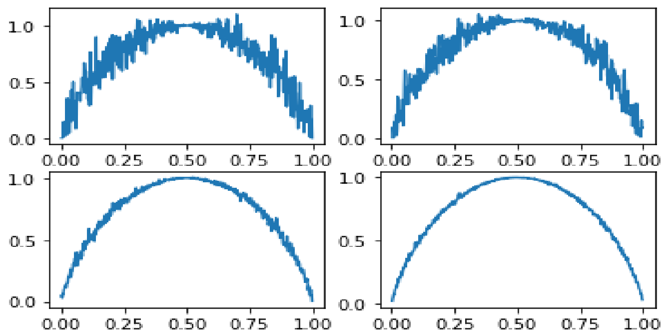

Figure 2: As N increases, the sample mean converges to Entropy

3 Typical Set Construction

We had 3 parameters namely p - the parameter of our bernoulli distribution, N- the length of the Sequence and ϵ, the convergence threshold. Based on the variation of these parameters, we tabulated our observations and made conclusions out of it.

Sequence Length	P = 0.1	P=0.2	P=0.25	P=0.3
8	8	17	28	84
9	9	36	36	120
10	10	45	165	120
11	11	55	165	495
12	12	272	220	715

Figure 3: The size of the typical set increases with N, and so is with the variation along P.

It can be seen that, as the length of the sequence increases, so is the number of **unique** elements in the typical set.
It can also be seen that as P increases, the entropy increases, hence the size of the typical set increases .

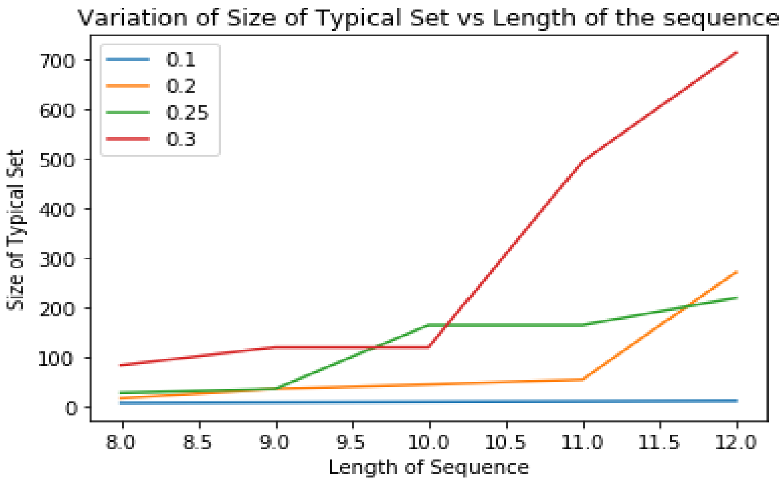

Figure 4: Variation of Size of Typical Set

4 Frequency Distribution of Typical Set

In this section, we are going to make the histogram plot to have a look at the frequency distribution of the Elements in the typical set. For this we are going to plot the decimal values of the Sequence vectors and hence will make their Histogram Plots.

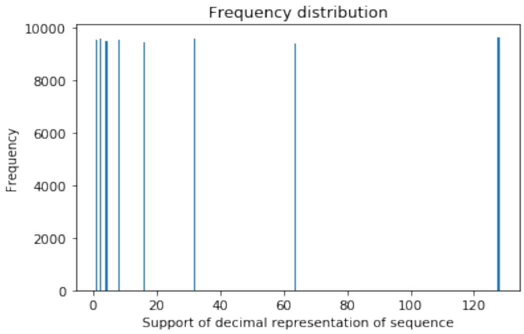

Figure 5: N=8, $\epsilon = 0.15$, P=0.1

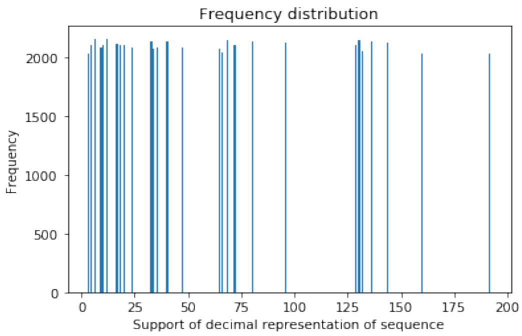

Figure 6: N=8, $\epsilon = 0.15$, P=0.2

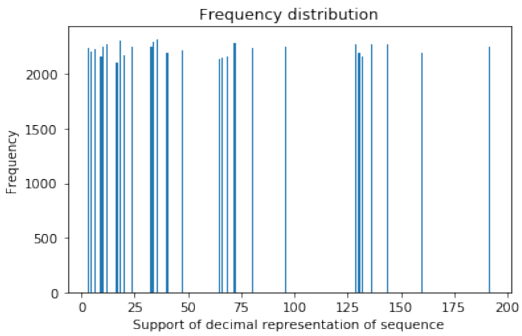

Figure 7: N=8, $\epsilon = 0.15$, P=0.25

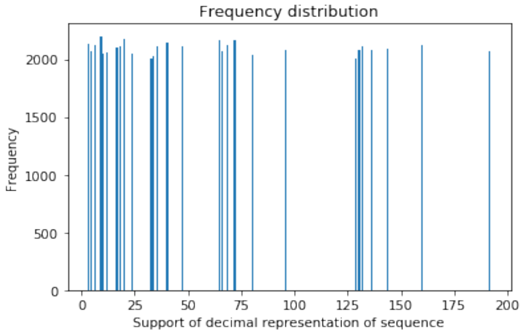

Figure 8: N=8, $\epsilon = 0.08$, P=0.3

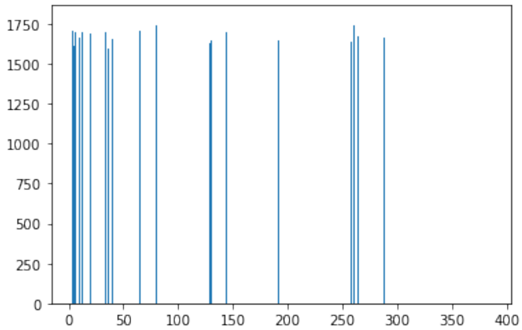

Figure 9: N=9, $\epsilon = 0.05$, P=0.2

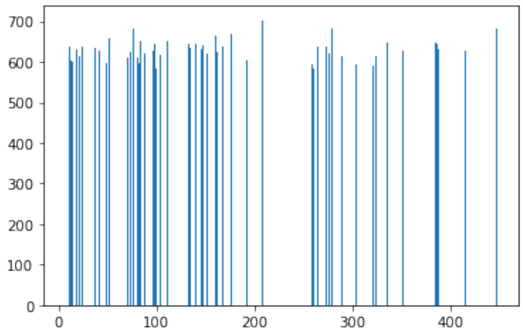

Figure 10: N=9, $\epsilon = 0.05$, P=0.3

5 Conclusions

- Thus from the frequency distribution graph, it was clear that the sequences falling inside the typical set make a more or less uniform distribution.
- As we increased the length of the sequence, the size of the typical set increased.
- As we increased the P value, the size of typical set increased.
- As we decrased the value of ϵ, the size of the typical set decreased.

Thus, it satisfies the most of postulates of the AEP.

6 Further Remarks and Acknowledgement

The source code can be extended to the source coding schemes and their study. I thank **Prof Jagdheesh Harshan** for the valuable topics that are covered by him in ELL714-Basic Information Theory. Code available as .ipynb file at following github **clickable** hyperlink.