ASYMPTOTIC EQUIPARTITION PROPERTY

Saurabh Kumar 2016EE10825,
Electrical Engineering Department, IIT Delhi
ee1160825@iitd.ac.in

***Abstract*— AEP states a very profound result, no matter how contradicting it may sound. It states that when observing a sequence of observations, as the number of observations increases, the probability of them lying in a particular ste also known as a typical set increases and reaches close to 1 for arbitrary larger value of observations. In this assignment we aim to explore this theorem and see the consequences of hyper parameters on the nature of this theorem. The hyperparamters are probability function, number of observations, the length of sequences to be observed and the bound on ϵ**

I. Definition of a Typical Set

Definition The typical set $A_\epsilon^{(n)}$ with respect to p(x) is the set of sequences $(x_1, x_2, \ldots, x_n) \in \mathcal{X}^n$ with the property

$$2^{-n(H(X)+\epsilon)} \leq p\left(x_1, x_2, \ldots, x_n\right) \leq 2^{-n(H(X)-\epsilon)} \tag{1}$$

As a consequence of the AEP, we can show that the set $A_\epsilon^{(n)}$ has the following properties:

Theorem :

1) If $(x_1, x_2, \ldots, x_n) \in A_\epsilon^{(n)}$ then $H(X) - \epsilon \leq -\frac{1}{n}\log p\left(x_1, x_2, \ldots\ x_n\right) \leq H(X) + \epsilon$
2) $Pr\left\{A_\epsilon^{(n)}\right\} > 1 - \epsilon$ for n sufficiently larger.
3) $\left|A_\epsilon^{(n)}\right| \leq 2^{n(H(X)+\epsilon)}$ where $|A|$ denotes the number of elements in the set A.
4) $\left|A_\epsilon^{(n)}\right| \geq (1-\epsilon)2^{n(H(X)-\epsilon)}$) for n sufficiently large.

Now, in this assignment we will try to see the consequences of varying different parameters affecting the AEP theorem.

II. Hyper Parameter Variation

Some of the various hyperparameter tuning and their consequences are mentioned in the next section. Without any other mentioning we take n =10 which is the length of the sequence we will observe, num of samples=10000 which is the total number of sequences we will be observing. Ideally this should be very high but since our computer can only handle a certain number of computations it is not feasible in the given time frame and resources to go for very high values of num of samples. We are going to observe a sequence of bernoulli random variable, so we need another parameter p to define that, which gives us another hyperparameter. By deault, p=0.5. Also the ϵ = 0.05, if not mentioned. So in a nutshell, if not mentioned,

- n =10
- num of samples= 10000
- p=0.5, and
- ϵ = 0.05

A. *Varying p rest constant*

The probability characteristic of any random variable is of utmost importance because it decides how the random variables are distributed. In this experiment we vary p to see its effect on the properties characterising the typical set.

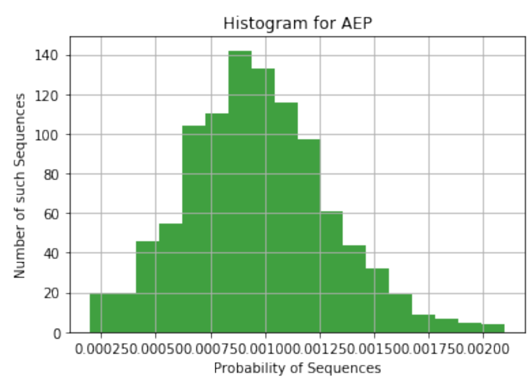

Fig. 1. Histogram of Sequence Probability and their corresponding frequency

- X-axis : Probability Ranges
- Y-axis : Number of Sequences lying in those probability ranges

This is the histogram for the default case where we have p=0.5. As we can see, we get a wide spectrum of values at the output. For n=10, we should get 1024 distinct sequences, and after running the code we indeed get **1024 distinct sequences**. The mean probability mass function of all the samples calculates was 0.0009765 with a variance of $1.0671e-07$. For a uniform bernoulli random variable with each entry independent of each other we should have pmf = 1/1024 = 0.0009765. This is same as what we got after simulation, hence the generation of samples is up to the mark.

*1) **Affect on Sequence Generation**:* Now we will try to observe what happens to the histogram as we vary p.

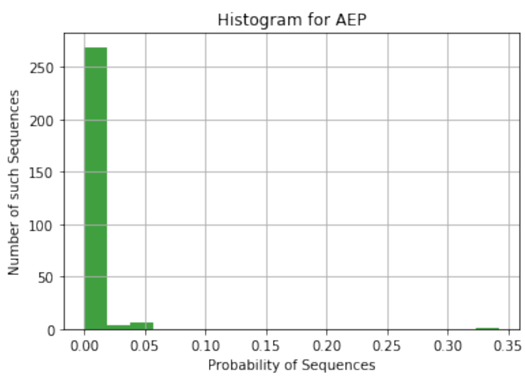

Fig. 2. Histogram of Sequence Probability and their corresponding frequency for p=0.1

- Total distinct sequences obtained = 276, contrary to 1024 for p=0.1
- Mean pmf = 0.003623, contrary to 0.0009765

This is also visualised from the histogram as we see most of the samples having probability in a very skewed range, and not all the samples were generated.

- Total distinct sequences obtained = 888, contrary to 1024 for p=0.3
- Mean pmf = 0.00112, contrary to 0.0009765

The data generated now has more samples, 888, compared to the case of p=0.1 and the pmf is also closer to the iid case.

- Total distinct sequences obtained = 1010, contrary to 1024 for p=0.4
- Mean pmf = 0.000990, contrary to 0.0009765

The data generated now has more samples, 1010, compared to the case of p=0.3 and the pmf is also closer to the iid case. Going with the same trend let's observe for p=0.7

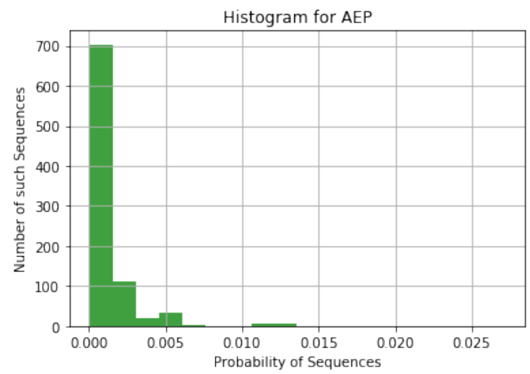

Fig. 3. Histogram of Sequence Probability and their corresponding frequency for p=0.3

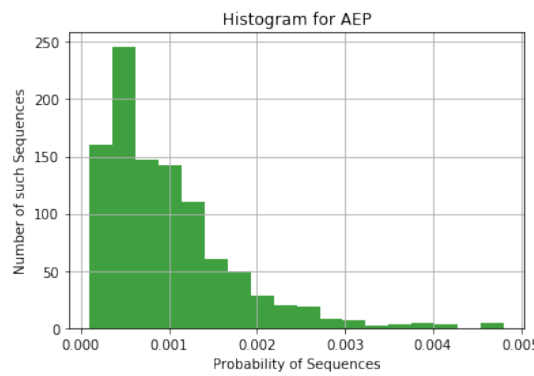

Fig. 4. Histogram of Sequence Probability and their corresponding frequency for p=0.4

- Total distinct sequences obtained = 877, contrary to 1024 for p=0.7
- Mean pmf = 0.001140, contrary to 0.0009765

So we observe this is again going down with the same trend it went up. Plotting this behaviour we get; So we see, as p increases upto 0.5, number of different sequences increases and reaches a max of 1024 and then decreases. This is because of the fact that on both of side of p=0.5 some sequences are more likely to be generated than others because at p=0.1 we are more likely to get a 0 and at p=0.9 we are more likely to get a 1.

*2) **Affect on Typical Set Construction** :* A typical set is defined by its elements, and its elements are a set of sequences which satisfy a certain set of mathematics axioms and inequalities. Now the question is to see how p, affects these axioms and hence the formation of typical set,

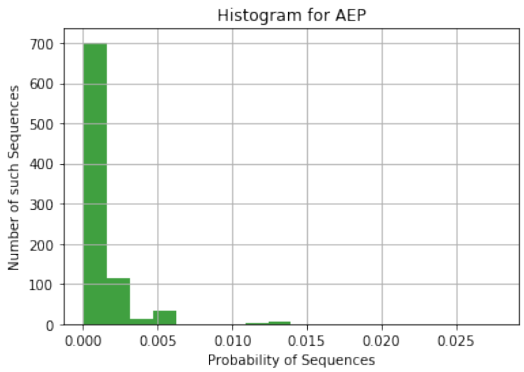

Fig. 5. Histogram of Sequence Probability and their corresponding frequency for p=0.7

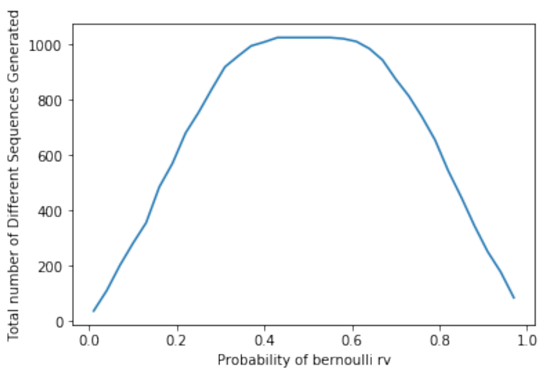

Fig. 6. Number of different sequence generate vs p

keeping rest constant. We observe the affect while varying num of samples too. We also want to verify that number of samples that we observe while making a typical set is also a leading factor. It is visually observed that the highest fraction of sequence the typical set has is close to 72% when p is close to 0.5. It is visually observed that the highest fraction of sequence the typical set has is close to 99% when p is close to 0.5.

Conclusion of this Experiment : As p increases the fraction of sequences belonging to the typical set increases and then decreases. When close to p=0.5, the AEP theorem is most likely to get satisfied for a given number of samples. Also as we increase the total number of observable samples, the $Pr\left\{A_{\epsilon}^{(n)}\right\} > 1 - \epsilon$ increases ans gets close to 1 as num of samples $\rightarrow \infty$.

B. Varying Epsilon ϵ , rest constant

We tried to study the variation of fraction of sequences falling under the typical set vs epsilon

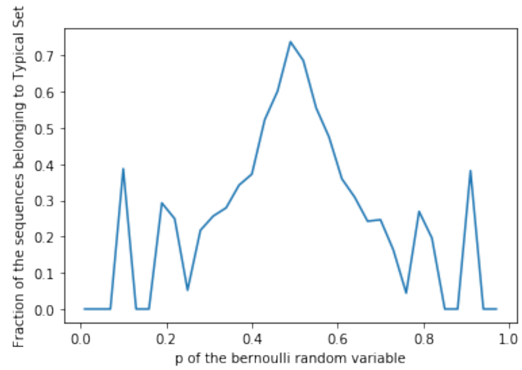

Fig. 7. Fraction of Sequences belonging to Typical Set vs p for num of samples = 10000

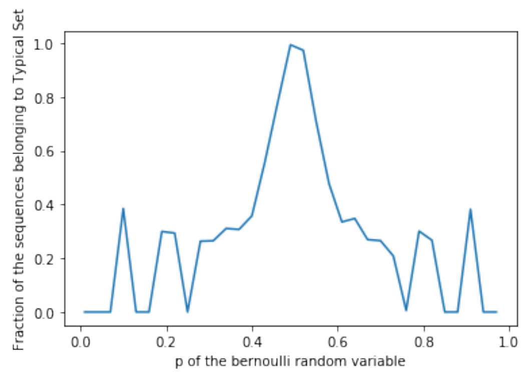

Fig. 8. Fraction of Sequences belonging to Typical Set vs p for num of samples = 100000

for two num of samples. Again to study parallel effects of epsilon as well of num of samples on deciding whether how many sequences will fall under the typical set. As epsilon value crosses a certain threshold, all the sequences automatically start to satisfy $H(X) - \epsilon \leq -\frac{1}{n} \log p\,(x_1, x_2, \ldots x_n) \leq H(X) + \epsilon$ and hence $Pr\left\{A_{\epsilon}^{(n)}\right\} > 1 - \epsilon$ holds for all of them.

As we increase the num of observable samples the graphs smoothens and ϵ threshold after which all the elements fall in typical set decreases too.

Conclusion of this Experiment : The bounds on epsilon surely affects if the bounds of a typical set will get satisfied or not. While a smaller ϵ makes it difficult for a sequence to fall inside a typical set, by increasing ϵ the bounds get relaxed and the sequence falls under typical set regime. So during our experiment we observed that for a given p, n and num of samples there exist a ϵ after which all the sequences fall under the typical set. We define

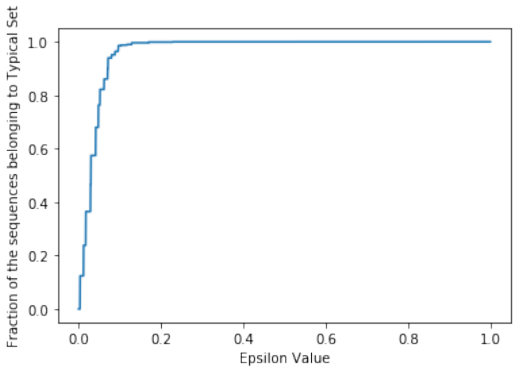

Fig. 9. Fraction of Sequences belonging to Typical Set vs ϵ for num of samples = 10000,p=0.5

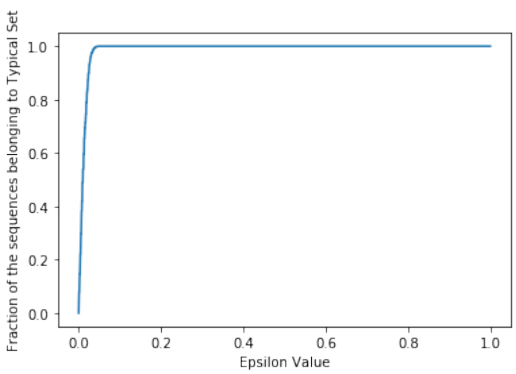

Fig. 10. Fraction of Sequences belonging to Typical Set vs ϵ for num of samples = 100000, p=0.5

that ϵ as **epsilon threshold**. Also one thing to note down is that as num of samples increased from 10000 to 100000, this threshold decreased, again showing the fact that by sampling a large pool of sequences of arbitrary lengths we can make any sequence fall under typical set with very high probability. Also as part of auxiliary observation when we change p to 0.25 the threshold increases from 0.1 to 0.7, representing now it is more difficult for all the sequences to fit inside the typical set bounds.

C. *Varying n, length of Sequence, rest constant*

One of the parts of AEP theorem states that as n increases, $Pr\left\{A_{\epsilon}^{(n)}\right\} > 1 - \epsilon$ this inequality gets satisfied. We will investigate this statement in this part. We will study how does the fraction of number of sequences falling under typical set vary when we vary n. And for larger num of observable samples; Let's also observe the number of distinct

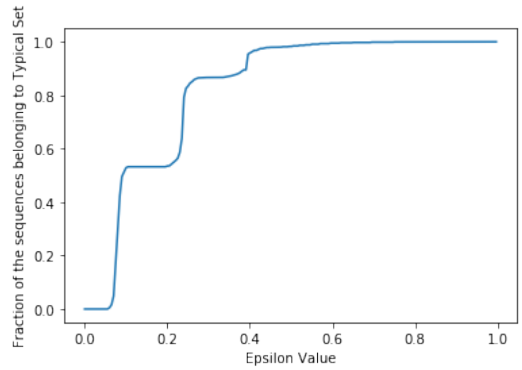

Fig. 11. Fraction of Sequences belonging to Typical Set vs ϵ for num of samples = 100000, p=0.25

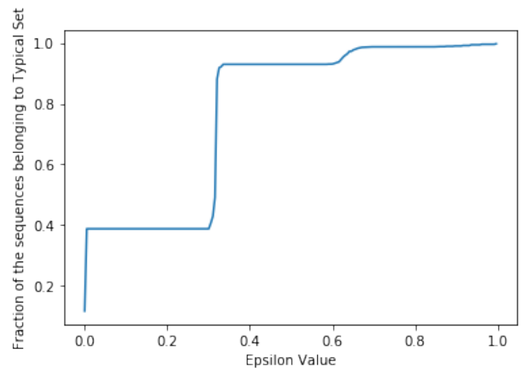

Fig. 12. Fraction of Sequences belonging to Typical Set vs ϵ for num of samples = 100000, p=0.90

sequences inside typical set also, as we vary n

This is very interesting to note. As we keep on increasing n, upto a limit the number of distinct sequences inside typical set increases, but after n=17 it starts decreasing. This has a **very meaning full reason** behind it. The total number of distinct sequences for n=17 is 2^{17}, which is more than 100000, hence a lot of such sequences won't even come once when generating sequences, and hence they won't lie inside the typical set either.

Conclusions of this experiment : Though it is possible for all n such that $Pr\left\{A_{\epsilon}^{(n)}\right\} > 1 - \epsilon$ holds, but in order to make sure this happens, as we increase n the number of observable have to increase too. Meaning for larger n we have to see more samples, then only most of them will fall under the typical set. This is straight forward from our two graphs. In one when num of samples =1e4, after a certain n, close to 14, the fraction of sequences going under typical set regimes falls

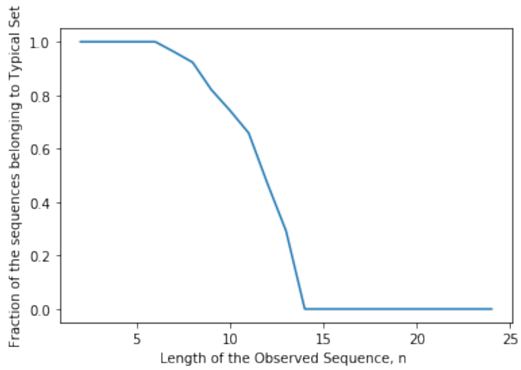

Fig. 13. Fraction of Sequences belonging to Typical Set vs n for num of samples = 10000

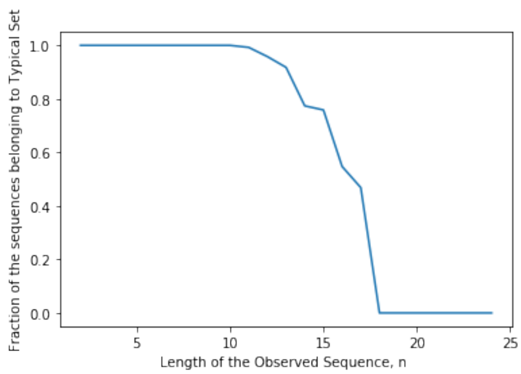

Fig. 14. Fraction of Sequences belonging to Typical Set vs n for num of samples = 100000

drastically but for num of samples = 1e5, this n increases, and the threshold shifts to n close to 18, meaning the earlier samples which were rejected for n close to 15 have certains chances of falling under typical set under larger number of observations. And as we'll continue this until infinity the threshold will increase and increase and we'll be able to accommodate all length sequences in typical set. But the takeaway point is, as n increases, the fraction of sequences falling under typical set for a particular p, and num of samples, decreases rapidly. Also we saw the implication of increasing n with respect to num of samples. We can only increase n upto a point where $2^n <$ **num of samples**, after that no matter how much we try, AEP won't hold for such n,p, ϵ and num of samples.

D. Varying num of samples, rest constant

From starting onwards we have put great emphasize on the fact that observing larger number of

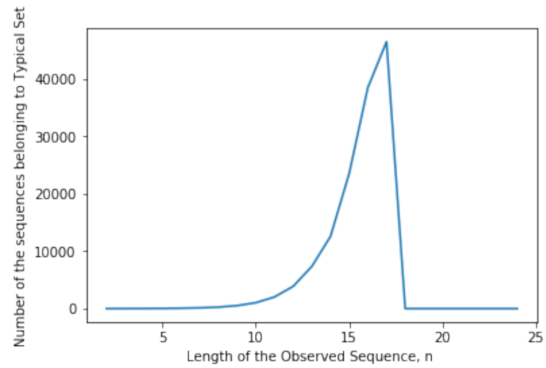

Fig. 15. Fraction of Sequences belonging to Typical Set vs n for num of samples = 100000

samples will give us better results and will increase the chances of sequences falling under the typical set. We'll see this in detail in this section. In this experiment we will vary the number of observed samples to see how much they affect the formation of a typical set. Also note that, we didn't go for very large values of num of samples because of the limited computation power and time in hand.

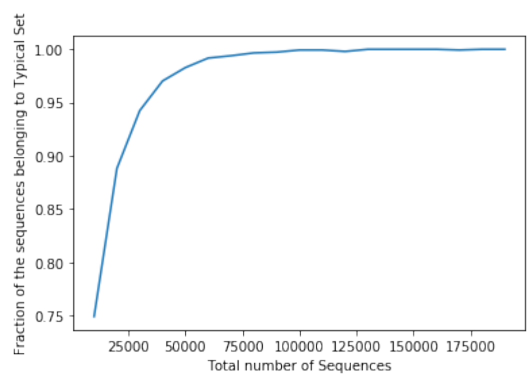

Fig. 16. Fraction of Sequences belonging to Typical Set vs num of observed samples

Conclusion of this Experiment: The trend is itself self explanatory. As we increase num of observed samples, the probability that the sequences lie inside typical set increases. And for very large values of num of samples, it reaches arbitrary close to 1. This verifies the AEP axioms and is in conjunction with law of large numbers too.

III. Overall Conclusions

AEP gives us a very crucial aspect of information theorem, that when observing a large number of observations, we only have to be concerned about few observable as they are the ones which will occur most frequently with high probabilities. In this assignment we discussed and simulated AEP and effect of its parameters such as n,p and num of sequences. We saw in detail how typical set behaves under all these parameter settings. We saw the importance of p while sampling. The more skewed the probabilities are, the worse the data distribution and hence lesser is the chances of all the sequences satisfying typical set probabilities condition. Then we saw the affects of ϵ bounds on typical set formation. Epsilon decides the bounds on whether a certain sequence should be a part of the typical set or not. Then we moved on to see the implications of increasing the length of observed sequences. And as a final summary we saw that increasing the number of observable always makes sure that typical set theorem holds better. One realization that we would like to point out is, in order to verify AEP we could have gone for two approaches, both correct. First being observing the total number of distinct sequences lying inside the typical set out of the 2^n total sequences, or to observe actually how many sequences out of the total observed sequences are lying in the typical set, both are correct methods of approach. We took the latter one for better understanding and for depicting more relationships. Here are some important points to note;

- For a given p,ϵ and num of samples, whenever $2^n<$ num of samples doesn't hold, the fraction of sequences belonging to typical set decreases rapidly and after that no matter how much we try we won't be able to include sequences in the typical set furthermore.
- ϵ decides the bound as to how muhc leverage we have to fit a sequence inside typical set. The more ϵ the more leverage we have.
- In essence num of samples are the most important aspect to make sure the condition of typical set holds after a certain number of observations. And practically as num of samples $\rightarrow \infty$ all sequences converges to typical set.

REFERENCES

[1] https://docs.scipy.org/doc/scipy-0.14.0/reference/generated/scipy.stats.bernoulli.html.

[2] ELEMENTS OF INFORMATION THEORY,THOMAS M. COVER JOY A. THOMAS.

[3] https://matplotlib.org/3.1.1/gallery/pyplots/pyplottext.htmlsphx-glr-gallery-pyplots-pyplot-text-py.

[4] https://matplotlib.org/3.1.0/api/asgen/matplotlib.pyplot.plot.html