Information Theory (ELL714)

Assignment 1

Asymptotic equipartition property

Submitted To:
Dr.Harshan Jagadish,
Assistant Professor,
Department of EE,
IIT-Delhi, India.

Submitted By :
Ajey Pai K
(2019EEY7534)

Contents

1   Objective

The aim of this assignment is to verify the asymptotic equipartition property and perform some observational experiments to obtain a good intuition of the concept.

2   Introduction

The asymptotic equipartition (AEP) theory is a general property of the observations of a stochastic process. The typical set, a consequential understanding of the AEP is most fundamental in theories of compression. The AEP is a direct consequence of the weak law of large numbers. Simply put, the AEP suggests that for any stationary, identically and independently distributed samples of a stochastic process,

$$\lim_{n \to \infty} Pr\left[|\frac{1}{n}\log(X_1, X2, X3, ..., X_n) - H(x)| > \epsilon\right] = 0 \ \forall \epsilon > 0 \qquad (1)$$

3   The experiments

1. Generate a bernoulli random variable with success probability p.
2. Generate a large set of sequences of such random variables.(with length n)
3. Plot a histogram of the large number of generated sequences.
4. Vary the value of p and observe any structure in the family of sequences thus generated.
5. Choose a value for ϵ and find what sequences lie in the typical set. Vary the value of n and observe what fraction of total sequences generated lie in the typical set.

4   Outputs and observations

Figure 1. shows the typical output terminal.


```
Length of sequence is: 8
No.of Sequences is: 50000
Bernoulli parameter p is: 0.2


The entropy H(X) is: 0.721928 bits
The size of the sample space is: 2 ^ 8
Total number of typical sequences are: 28
The fraction of sequences in the typical set is: 0.113360323886
```

Figure 1: Output for p = 0.2, ϵ = 0.1, n=8, m=50000

For this combination of inputs, the histogram of all sequences obtained is shown in figure 2.
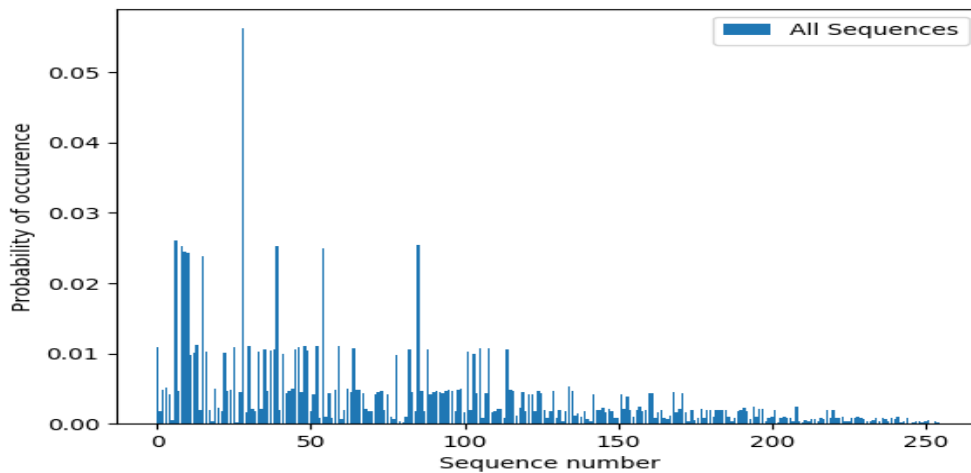
(1).png (1).png

Figure 2: Histogram of all generated sequences

When p was varied for a fixed length of sequences and some ϵ, the following results were obtained. Figure 3 shows the histograms of all sequence generated and the corresponding typical sets are shown in Figure 3 through Figure 6. The major observation is that with any value of p, it is certain that majority of the typical sequences occur earlier during the generation. In the following plots, n=8 and m = 50000 (No. of sequences totally generated). Similarly, the observation plots
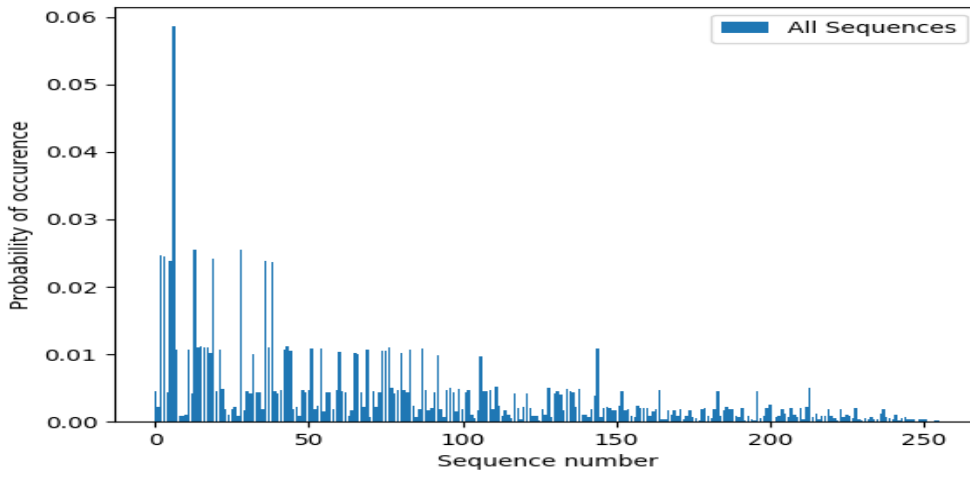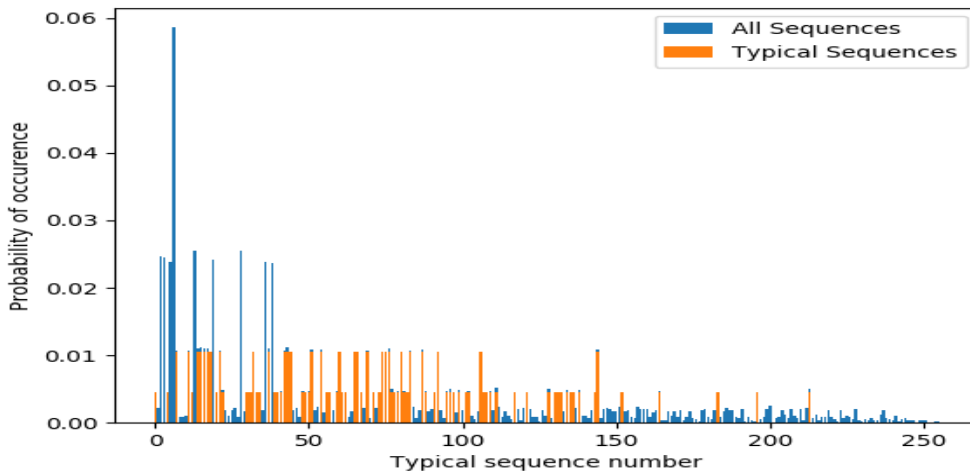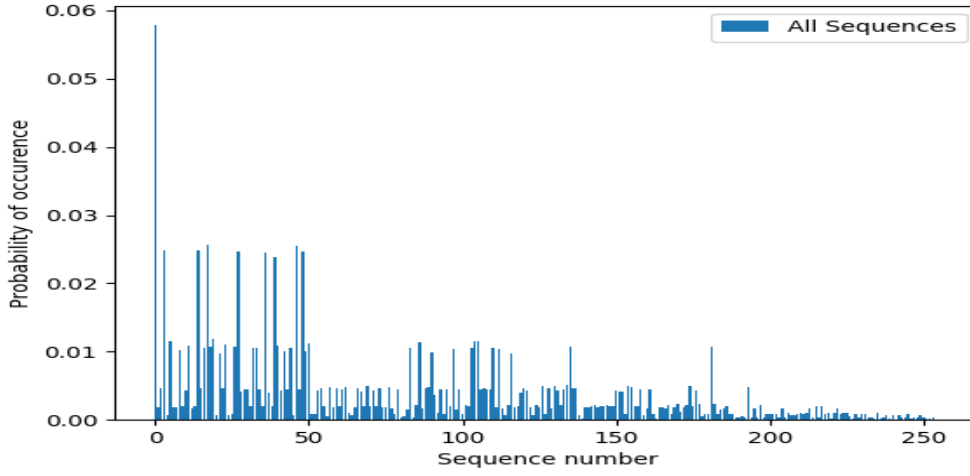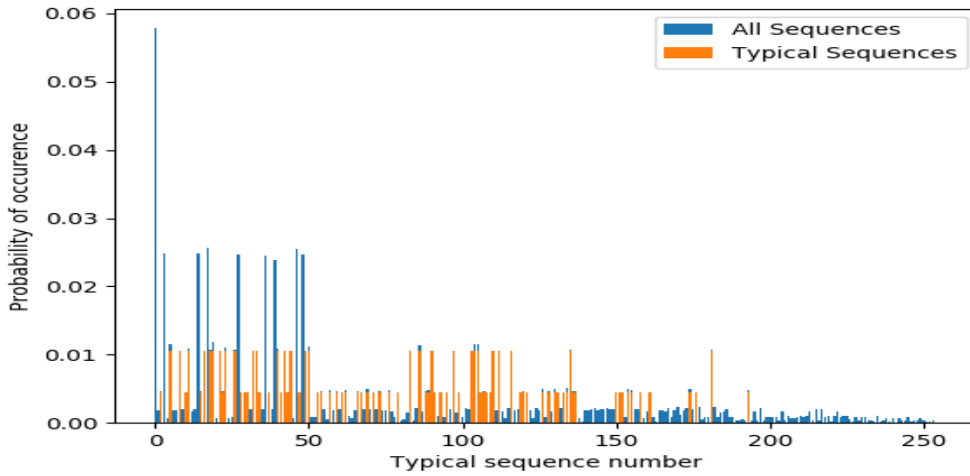

Figure 3: Histogram of sequences with p=0.3

Figure 4: Histogram of sequences with p=0.3

Figure 5: Histogram of sequences with p=0.7

Figure 6: Histogram of sequences with p=0.7

for when n was changed is shown in Figures 7 through 10. While choosing an n such that (2,n) ¡ m(total sequences generated), then a certain percentage of the generated sequence will lie in the typical set and not all of them. Conversely, while choosing n such that (2,n) ¿ m (Total no. Of sequences generated), then all the generated sequence will lie in the typical set. In the following figures, p=0.7, m = 50000.
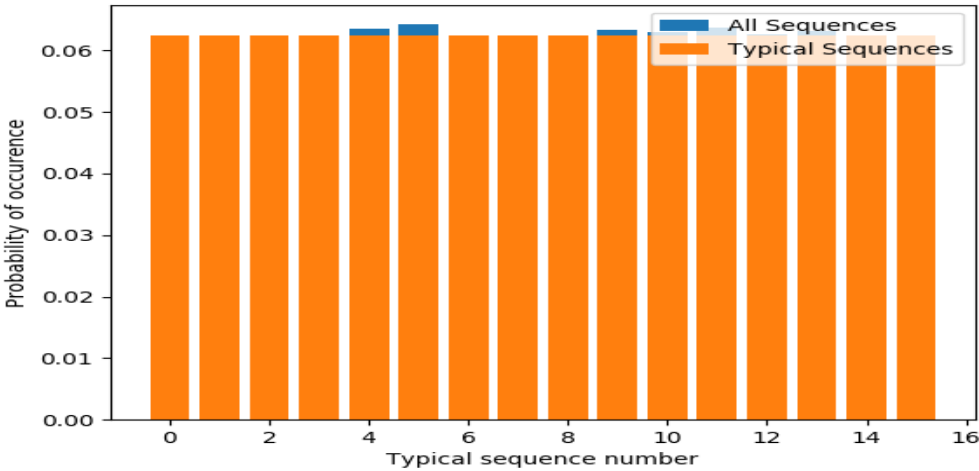
Figure 7: Histogram of sequences with n=4

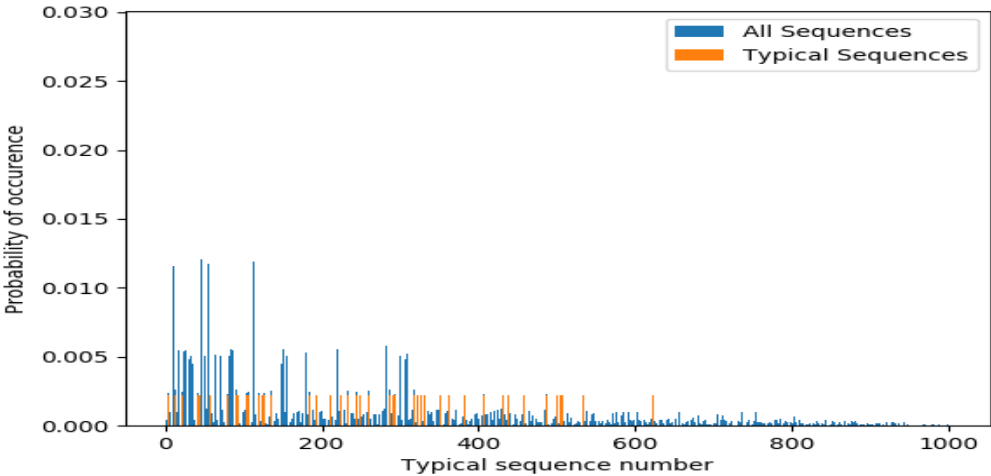Figure 8: Histogram of sequences with n=10