A Different Approach to Asymptotic Equipartition Property (AEP)

Neelabh Keshav[1], Dr. Bikas Chandra Kumar[2]

Student-M.Tech. (Communication System Engg.), IIT Patna[1], Deputy Director (Comp.), Dept. of Sc. & Tech., Govt. of Bihar[2]

Abstract

The AEP theorem [1-2] predicts occurrence of a typical set of length n, such that n is a very large number. These typical sets are very small in number compared to the total number of different sets possible with n number of symbols (= m^n, where m is the number of different symbols from DMS). Although very small, it is this typical set which has very high combined probability of occurrence (≈1). This typical set is nothing but different arrangement of symbols each of whose number is governed by their probability of occurrence. Source Encoding Theorem [1] involves encoding these different typical sets. In case, atypical set occurs, there is no coding for it. The source coding theorem is then said to fail, although probability of failure is very low.

Keywords: typical set, atypical set, DMS, combined probability, Source coding theorem.

1. Law of large numbers

Let an un-biased coin be taken. If this coin is tossed, one will either get heads or tail. Now let the coin be tossed and start keeping record of outcome each time. In case, coin is tossed for n number of times such that n is very large number, it is known by experience, there will be an equal number of heads and tail. This may not be the case if n is small. If same experiment is performed with a dice, all the outcomes will be equal in number. In case of coin toss 2^n numbers of different outcomes are possible. Same applies to dice, where 6^n numbers of different outcomes are possible. In spite of many possible outcomes, with n→∞, the only thing that can happen, as one goes by his experience is equal number of heads and tails in case of coin toss for unbiased coin and equal number of 1, 2, 3, 4, 5 and 6 in case of dice throw for unbiased dice.

2. The AEP (Asymptotic Equipartition Property)

Let there be a DMS, output of DMS Y_j such that, Y_1, Y_2, Y_3…….. are a sequence of iid (independent and identically distributed) random variables.

Now, sample average,

$$A_Y^n = \frac{Y_1 + Y_2 + \ldots\ldots + Y_n}{n} = \frac{S_Y^n}{n} \quad \text{(eq.1)}$$

A_Y^n is a r.v., so is S_Y^n.

Mean of S_Y^n is thus $n\overline{Y}$. Variance of S_Y^n is $n\sigma_Y^2$

Similarly, Mean of A_Y^n is thus $\overline{Y}$. Variance of A_Y^n is $\frac{\sigma_Y^2}{n}$

Now, from Chebyshev inequality,

$$\Pr[\,|A_Y^n - \overline{Y}| \geq \varepsilon\,] \leq \frac{\sigma_Y^2}{n\varepsilon^2} \quad \text{(eq.2)}$$

For any ε>0, and n→∞,

$$\Pr[\,|A_Y^n - \overline{Y}| \geq \varepsilon\,] \rightarrow 0 \quad \text{(eq.3)}$$

So, it means, if one take a large number of iid, find sample average, find probability of r.v., beyond a window ε (+ve), is 0. So, all sample average fall on mean.

Let a term be defined as,

$$W(x) = -\log_2 p_X(x) \quad \text{(eq.4)}$$

$W(X_k)$ gives value $W(x)$ when $x = X_k$. This $W(x)$ will be random in nature, and its probability of occurrence Pr $[W(x)] = \Pr(x)$, as $W(x)$ is a one to one function of x. Mean of $W(x)$,

$$E[W(X_k)] = \sum_{j=1}^{j=m}[\text{pmf of } W(X_k)][W(X_k)]$$

$$= \sum_{j=1}^{j=m}[p_X(x_j)][W(X_k)]$$

$$= \sum_{j=1}^{j=m} -p_X(x_j)\log_2 p_X(x)$$

$$= H[X] \qquad \text{(eq.5)}$$

Again, let a chunk of symbol with length n be defined.

$X^n = \{X_1, X_2, \ldots \ldots X_n\}$. Each of this X_j, can be anything from the set x_j which is m in number.

Now, sample average is,

$$A_w^n = \frac{W(X_1) + \ldots \ldots W(X_n)}{n}$$

$$= \frac{-\log_2 p_X^n(x^n)}{n} \qquad \text{(eq.6)}$$

Now, from chebyshev inequality,

$$\Pr[\,|A_w^n - \overline{A_w^n}| \ge \varepsilon\,] \le \frac{\sigma_w^2}{n\varepsilon^2} \qquad \text{(eq.7)}$$

σ_w^2 is variance of $W(X_k)$, $1 \le k \le \infty$

Now,

$$\Pr[\,|A_w^n - E[A_w^n]| \ge \varepsilon\,] \le \frac{\sigma_w^2}{n\varepsilon^2}$$

$$\Rightarrow \Pr\left[\,\left|\frac{-\log_2 p_X^n(x^n)}{n} - H[X]\right| \ge \varepsilon\,\right] \le \frac{\sigma_w^2}{n\varepsilon^2} \qquad \text{(eq.8)}$$

Define a typical set such that,

$$T_\varepsilon^n = \left\{ x^n : \left|\frac{-\log_2 p_X^n(x^n)}{n} - H[X]\right| \le \varepsilon \right\} \qquad \text{(eq.9)}$$

Thus, probability of typical set is given by

$$\Pr \ge 1 - \frac{\sigma_w^2}{n\varepsilon^2} \qquad \text{(eq.10)}$$

(For any $\varepsilon > 0$)

For n→∞, i.e., very large size of symbol blocks, Pr of typical set→1. Now,

$$T_\varepsilon^n = \left\{ x^n : \left|\frac{-\log_2 p_X^n(x^n)}{n} - H[X]\right| \le \varepsilon \right\}$$

$$= \{ x^n : |-\log_2 p_X^n(x^n) - nH[X]| \le n\varepsilon \}$$

$$= \{ x^n : -n\varepsilon \le -\log_2 p_X^n(x^n) - nH[X] \le n\varepsilon \}$$

$$= \{ x^n : 2^{-nH[X]+n\varepsilon} > p_X^n(x^n) > 2^{-nH[X]-n\varepsilon} \} \qquad \text{(eq. 11)}$$

Suppose $nH[X] \gg n\varepsilon$, the probability of each member in typical set becomes,

$$p_X^n(x^n) \approx 2^{-nH[X]} \qquad \text{(eq.12)}$$

Thus, for a large n, computed joint probability will always be equal to this equation.

$$p_X^n(x^n) > 2^{-nH[X]-n\varepsilon}$$

$$\text{and,} \quad 1 \ge \sum p_X^n(x^n) > 2^{-nH[X]-n\varepsilon} |T_\varepsilon^n|$$

$$\text{so,} \quad |T_\varepsilon^n| < 2^{nH[X]+n\varepsilon} \qquad \text{(eq.13)}$$

where,

$|T_\varepsilon^n|$ denotes number of typical sets.

also,

$$1 - \frac{\sigma_w^2}{n\varepsilon^2} \le \Pr(T_\varepsilon^n)$$

$$\Rightarrow 1 - \frac{\sigma_w^2}{n\varepsilon^2} \le p_X^n(x^n) < |T_\varepsilon^n|\, 2^{-(nH[X]-n\varepsilon)}$$

$$\Rightarrow |T_\varepsilon^n| > \left(1 - \frac{\sigma_w^2}{n\varepsilon^2}\right) 2^{(nH[X]-n\varepsilon)}$$

Therefore, $(1-\delta)\,2^{n(H[x]-\varepsilon)} < |T_\varepsilon^n| < 2^{n(H[x]+\varepsilon)}$ (eq.14)

Again, for $nH[X] \gg n\varepsilon$,

$$|T_\varepsilon^n| \approx 2^{nH[x]} \qquad \text{(eq.15)}$$

Aggregated probability,

$$p_X^n(x^n) \; |T_\varepsilon^n| \approx 2^{n\,H[x]} \; 2^{-nH[X]} \approx 1 \qquad \text{(eq.16)}$$

Moreover, with n blocks of m different symbols, m^n symbol sets are possible. However, above expression predicts that typical set has a probability of occurrence of almost 1. So out of these m^n possible sets, only $2^{nH[X]}$ sets of symbols are possible.

$$\frac{m^n}{2^{nH[X]}} = \frac{2^{n\log_2 m}}{2^{nH[X]}} = 2^{n(\log_2 m - H[X])} \qquad \text{(eq.17)}$$

For any DMS, one knows that, $H[X] < \log_2 m$. So, $(\log_2 m - H[X])$ is always positive, and as $n \to \infty$, the above ratio also tends to ∞.

From these mathematical operations, one can infer that, if he has a DMS and take an n block of symbols from this DMS, such that n is very large, he'll find that it is the typical set (which although very small in number), has a very high probability of occurrence.

3. Experience vs. AEP

If one goes by experience, an equal number of all events in an equiprobable situation will be obtained. In general, if there are finite sets of elements in sample space, such that each have their respective probability of occurrence. In that case, if a large number of samples (=n) from a DMS are taken, this set will contain the different elements, such that the ratio of their number to total n will be equal to their probability of occurrence. In other words the number will be governed by the cdf of the elements.

However, by mathematical calculations, it was found that there is high probability of occurrence of typical set, where

$$(1 - \delta)\, 2^{n\,(H[x] - \varepsilon)} < \; |T_\varepsilon^n| \; < 2^{n\,(H[x] + \varepsilon)} \qquad \text{(eq.18)}$$

and, probability of occurrence of this typical set is given by,

$$\Pr(T_\varepsilon^n) \geq 1 - \frac{\sigma_w^2}{n\,\varepsilon^2} \qquad \text{(eq.19)}$$

which tends to 1 when $n \to \infty$. This means that a DMS will emit samples which when taken in blocks of n such that n is very large, it will almost always be a typical set whose number has been specified.

So, either one of the two ways is wrong or both of them point to the same thing. In fact, the latter is true, i.e.,

when samples of DMS are taken in blocks of n ($n \to \infty$), almost every time this set is typical set. This typical set is set of symbols, such that their number of occurrence in block of n is governed by their probability of occurrence. Almost every time their number will be as per their probability distribution, but their arrangement will differ, such that the number of typical sets will be given by equation (18).

4. Mathematical Verification

$$p_X(x^n) \approx 2^{-nH[X]}$$
$$=> \log_2 p_X(x^n) \approx -nH[X]$$

Thus,

$$H[X] \approx \frac{1}{n}\log_2 \frac{1}{p_X^n(x^n)} \qquad \text{(eq.20)}$$

So, this typical set having n blocks of symbols must satisfy the equation (20).

MATLAB was used to generate random numbers from 1 to 8; each number representing a symbol from DMS and unequal probability of occurrence of symbols was used taking blocks of 800 and 80,000 ; P(1)= 0.015625; P(2)=0.078125; P(3)=0.078125; P(4)=0.265625; P(5)=0.265625; P(6)=0.015625; P(7)=0.140625; and P(8)=0.140625, where P(X) is probability of occurrence of symbol X. H[X] of DMS with this P[X] = 2.5742.

From the fig.1 and fig.2 which represents the number of different symbols in one block of n, the value for following expression was found,
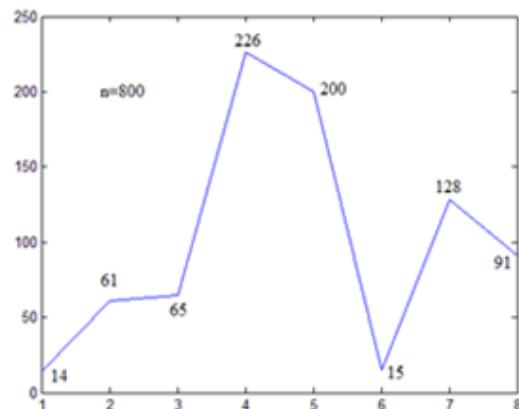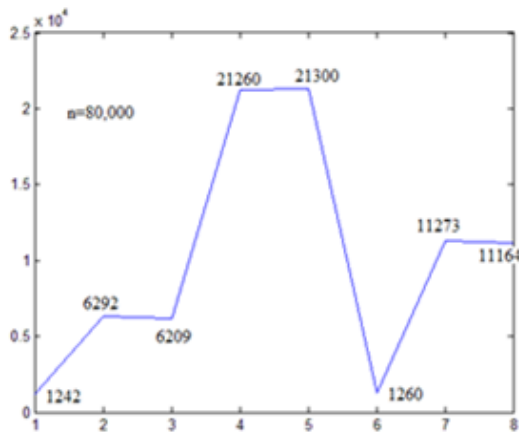
$$\frac{1}{n}\log_2 \frac{1}{p_X(x^n)}$$

Fig.1

Fig.2

[1] Robert Gallager, course materials for 6.450 Principles of Digital Communications I, Fall 2006. MIT OpenCourseWare (http://ocw.mit.edu/), Massachusetts Institute of Technology, Pg. 36-41.

[2] Lecture 5: *Asymptotic Equipartition Property* - Duke University (*people.ee.duke.edu/~yx44/ece587/Lecture5.pdf*).

Now for fig 1, value = 2.5899; and 2.57355 for fig.2.

The above equation, in general, can be applied to any probability distribution and any number of finite sets, with the equality approaching as $n \to \infty$, as can be seen in the above results where the value is closer to H[X] in second case which has greater number of n.

5. Conclusion

AEP predicts an existence of "typical sets" of n symbols which is very small in number compared to m^n number of total possible sets, where m is the number of different symbols, but has very high probability of occurrence ≈ 1. This typical set is nothing but different possible arrangement of m symbols in n blocks, where the number of each symbol is governed by its probability of occurrence. Thus, by AEP it was actually proved what one knows by experience such as if he keep on tossing the coin and maintain the record of outcome, a large number of samples will have half number of tails and half number of heads for an unbiased coin.

6. Acknowledgement

7. References