INFORMATION THEORY & CODING

Part 4 : Asymptotic Equipartition Property (AEP)

Dr. Rui Wang

Department of Electrical and Electronic Engineering
Southern Univ. of Science and Technology (SUSTech)

Email: wang.r@sustech.edu.cn

October 10, 2023

Stock Market

- Initial investment Y_0, daily return ratio r_i, in t-th day, your money is

$$Y_t = Y_0 r_1 \cdot \ldots \cdot r_t.$$

- Now if returns ratio r_i are i.i.d., with

$$r_i = \begin{cases} 4, & \text{w.p. } 1/2 \\ 0, & \text{w.p. } 1/2 \end{cases}$$

- So you think the expected return ratio is $E[r_i] = 2$.
- And then

$$E[Y_t] = E[Y_0 r_1 \cdot \ldots \cdot r_t] = Y_0 (E[r_i])^t = Y_0 2^t ???$$

Stock Market

- With $Y_0 = 1$, actual return Y_t goes like

$$1 \quad 4 \quad 16 \quad 0 \quad 0 \quad \cdots$$

- Why?
  - The 'typical' sequences will end up with 0 return.
  - Occasionally, we got high return.
  - The expected return is increasing.
  - Expectation does not show the typical feature of this random sequence. We can turn to typical set.

Weak Law of Large Numbers

Theorem (Weak Law of Large Numbers)

Suppose that $X_1, X_2, \ldots, X_n$ are n independent, identically distributed (i.i.d.) random variables, then

$$\frac{1}{n}\sum_{i=1}^{n} X_i \to E[X] \qquad \text{in probability},$$

i.e. for every number $\epsilon > 0$,

$$\lim_{n\to\infty} \Pr\left[|\frac{1}{n}\sum_{i=1}^{n} X_i - E[X]| \leq \epsilon \right] = 1.$$

Asymptotic Equipartition Property (AEP)

Definition (Convergence of random variables)

Given a sequence of random variables, $X_1, X_2, \ldots,$ we say that the sequence $X_1, X_2, \ldots$ converges to a random variable X:

1. In probability if for every $\epsilon > 0$, $\Pr[|X_n - X| \geq \epsilon] \to 0$
2. In mean square if $E[(X_n - X)^2] \to 0$
3. With probability 1 (a.k.a. almost surely) if $\Pr[\lim_{n\to\infty} X_n = X] = 1$

Asymptotic Equipartition Property (AEP)

Theorem 3.1.1 (AEP)

If $X_1, X_2, \ldots$ are i.i.d. $\sim p(x)$, then

$$-\frac{1}{n}\log p(X_1, X_2, \ldots, X_n) \to H(X) \qquad \textit{in probability}.$$

Proof.

Since X_i are i.i.d., so are $\log p(X_i)$. Hence, by the weak law of large numbers,

$$\begin{aligned}
-\frac{1}{n}\log p\left(X_1, X_2, \ldots, X_n\right) &= -\frac{1}{n}\sum_i \log p\left(X_i\right) \\
&\to -E[\log p(X)] \qquad \text{in probability} \\
&= H(X)
\end{aligned}$$

$\square$

Typical Set

Definition

A *typical set* $A_\epsilon^{(n)}$ contains all sequence realizations $(x_1, x_2, \ldots, x_n) \in \mathcal{X}^n$ with

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

Consequences of AEP

Theorem 3.1.2

- *If* $(x_1, x_2, \ldots, x_n) \in A_\epsilon^{(n)}$, *then* $H(X) - \epsilon \le -\frac{1}{n}\log p(x_1, x_2, \ldots, x_n) \le H(X) + \epsilon.$
- $\Pr[(X_1, X_2, ..., X_n) \in A_\epsilon^{(n)}] > 1 - \epsilon$ *for n sufficiently large.*
- $|A_\epsilon^{(n)}| \le 2^{n(H(X)+\epsilon)}$, *where* $|A|$ *denotes the cardinality of the set* A.
- $|A_\epsilon^{(n)}| \ge (1-\epsilon)2^{n(H(X)-\epsilon)}$ *for n sufficiently large.*

Proof.

1. Immediate from the definition of $A_\epsilon^{(n)}$. □

The number of bits used to describe sequences in typical set is approximately $nH(X)$.

Consequences of AEP

Theorem 3.1.2

- *If* $(x_1, x_2, \ldots, x_n) \in A_\epsilon^{(n)}$, *then* $H(X) - \epsilon \leq -\frac{1}{n} \log p(x_1, x_2, \ldots, x_n) \leq H(X) + \epsilon$.
- $\Pr[A_\epsilon^{(n)}] > 1 - \epsilon$ *for n sufficiently large.*
- $|A_\epsilon^{(n)}| \leq 2^{n(H(X)+\epsilon)}$, *where* $|A|$ *denotes the cardinality of the set* A.
- $|A_\epsilon^{(n)}| \geq (1-\epsilon)2^{n(H(X)-\epsilon)}$ *for n sufficiently large.*

Proof.

2. By Theorem 3.1.1, the probability of the event $(X_1, X_2, \ldots, X_n) \in A_\epsilon^{(n)}$ tends to 1 as $n \to \infty$. Thus, for any $\delta > 0$, there exists an n_0 such that for all $n \geq n_0$, we have

$$\Pr\left\{\left|-\frac{1}{n}\log p\left(X_1, X_2, \ldots, X_n\right) - H(X)\right| < \epsilon\right\} > 1 - \delta.$$

Setting $\delta = \epsilon$, the conclusion follows. $\square$

Consequences of AEP

Theorem 3.1.2

- *If* $(x_1, x_2, \ldots, x_n) \in A_\epsilon^{(n)}$, *then* $H(X) - \epsilon \leq -\frac{1}{n} \log p(x_1, x_2, \ldots, x_n) \leq H(X) + \epsilon$.
- $\Pr[A_\epsilon^{(n)}] > 1 - \epsilon$ *for n sufficiently large.*
- $|A_\epsilon^{(n)}| \leq 2^{n(H(X)+\epsilon)}$, *where $|A|$ denotes the cardinality of the set A.*
- $|A_\epsilon^{(n)}| \geq (1 - \epsilon)2^{n(H(X)-\epsilon)}$ *for n sufficiently large.*

Proof.

3.

$$\begin{aligned} 1 = \sum_{\mathbf{x} \in \mathcal{X}^n} p(\mathbf{x}) &\geq \sum_{\mathbf{x} \in A_\epsilon^{(n)}} p(\mathbf{x}) \\ &\geq \sum_{\mathbf{x} \in A_\epsilon^{(n)}} 2^{-n(H(X)+\epsilon)} \\ &= 2^{-n(H(X)+\epsilon)} \left| A_\epsilon^{(n)} \right|. \end{aligned}$$

□

Consequences of AEP

Theorem 3.1.2

- *If* $(x_1, x_2, \ldots, x_n) \in A_\epsilon^{(n)}$, *then* $H(X) - \epsilon \le -\frac{1}{n}\log p(x_1, x_2, \ldots, x_n) \le H(X) + \epsilon$.
- $\Pr[A_\epsilon^{(n)}] > 1 - \epsilon$ *for* n *sufficiently large.*
- $|A_\epsilon^{(n)}| \le 2^{n(H(X)+\epsilon)}$, *where* $|A|$ *denotes the cardinality of the set* A.
- $|A_\epsilon^{(n)}| \ge (1-\epsilon)2^{n(H(X)-\epsilon)}$ *for* n *sufficiently large.*

Proof.

4. For sufficiently large n, $\Pr[A_\epsilon^{(n)}] > 1 - \epsilon$, so that

$$\begin{aligned} 1 - \epsilon &< \Pr\left[A_\epsilon^{(n)}\right] \\ &\le \sum_{\mathbf{x} \in A_\epsilon^{(n)}} 2^{-n(H(X)-\epsilon)} \\ &= 2^{-n(H(X)-\epsilon)} \left|A_\epsilon^{(n)}\right|. \end{aligned}$$

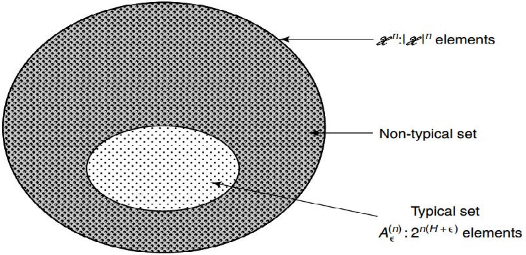
Typical set diagram

This enables us to divide all sequences into two sets

- Typical set: high probability to occur, sample entropy is close to true entropy

  so we will focus on analyzing sequences in typical set

- Non-typical set: small probability, can ignore in general



Theorem 3.2.1

Let $X_1, X_2, ..., X_n$ be i.i.d. random variables with distribution $p(x)$, and $X^n = X_1X_2...X_n$. For arbitrarily small $\epsilon > 0$, there exists a code that maps every realization $x^n = x_1x_2...x_n$ of X^n into one binary string, such that the mapping is one-to-one (and therefore invertible) and

$$E\left[\frac{1}{n}\ell(X^n)\right] \leq H(X) + \epsilon$$

for a sufficiently large n.

Asymptotic Equipartition Property (AEP)

Theorem 3.2.1

$$E\left[\frac{1}{n}\ell(X^n)\right] \leq H(X) + \epsilon.$$

for n sufficiently large.

Proof.

Description in typical set requires no more than $n(H(X) + \epsilon) + 1$ bits (correction of 1 bit because of integrality).

Description in atypical set ${A_\epsilon^{(n)}}^C$ requires no more than $n \log |\mathcal{X}| + 1$ bits.

Add another bit to indicate whether in $A_\epsilon^{(n)}$ or not to get whole description. □

Asymptotic Equipartition Property (AEP)

Theorem 3.2.1

$$E[\frac{1}{n}\ell(X^n)] \leq H(X) + \epsilon.$$

for n sufficiently large.

Proof.

Let $\ell(x^n)$ be the length of the binary description of x^n. Then, $\forall \epsilon > 0$, there exists n_0 s.t. $\forall n > n_0$,

$$\begin{aligned}
E\left(\ell\left(X^n\right)\right) =& \sum_{x^n} p\left(x^n\right)\ell\left(x^n\right) \\
=& \sum_{x^n \in A_\epsilon^{(n)}} p\left(x^n\right)\ell\left(x^n\right) + \sum_{x^n \in A_\epsilon^{(n)C}} p\left(x^n\right)\ell\left(x^n\right) \\
\leq& \sum_{x^n \leq A_\epsilon^{(n)}} p\left(x^n\right)\left(n(H+\epsilon)+2\right) + \sum_{x^n \in A_\epsilon^{(n)C}} p\left(x^n\right)\left(n\log|\mathcal{X}|+2\right) \\
=& \Pr[A_\epsilon^{(n)}](n(H+\epsilon)+2) + \Pr[A_\epsilon^{(n)^C}](n\log|\mathcal{X}|+2) \\
\leq& n(H+\epsilon) + \epsilon n(\log|\mathcal{X}|) + 2 \\
=& n\left(H+\epsilon'\right)
\end{aligned}$$

where $\epsilon' = \epsilon + \epsilon\log|\mathcal{X}| + \frac{2}{n}$ can be made arbitrarily small by choosing n properly. $\square$

Reading

Reading : Whole Chapter 3