# INFORMATION THEORY & CODING

## Week 4 : Asymptotic Equipartition Property (AEP)

(渐近等分性)

Dr. Rui Wang

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

Email: wang.r@sustech.edu.cn

October 11, 2022

# 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

$$1 \qquad 4 \qquad 16 \qquad 0 \qquad 0 \qquad \cdots$$

- 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 (大数定律)

## Theorem (Weak Law of Large Numbers)

(独立同分布)

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

概率渐近收敛于

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

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

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

# 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 $\Pr[\lim_{n\to\infty} X_n = X] = 1$

从这开始讲.

# Asymptotic Equipartition Property (AEP)

$\forall \varepsilon > 0, \lim_{n\to\infty} \Pr\left[ \left| -\frac{1}{n}\log p(X_1 \cdots X_n) - H(X) \right| \le \varepsilon \right] = 1.$

## Theorem 3.1.1 (AEP)

joint distribution.

*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,

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

□

# 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

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

$$-\epsilon \leq -\frac{1}{n}\log p(x_1, x_2, \cdots x_n) - H(X) \leq \epsilon$$

$$H(X)-\epsilon \leq -\frac{1}{n}\log p(x_1, x_2, \cdots x_n) \leq H(X)+\epsilon$$

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

$\rightarrow p(x_1, x_2, \cdots, x_n) \approx 2^{-nH(X)} \rightarrow$ 解释了equipartition.

$$A_\epsilon^{(n)} = \left\{(x_1, \cdots x_n) \;\middle|\; \left|-\frac{1}{n}\log p(x_1, x_2, \cdots, x_n) - H(X)\right| \leq \epsilon\right\}$$

## 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 $nH(X)$.

$\forall \epsilon > 0, \lim_{n\to\infty} \Pr[|-\frac{1}{n}\log p(X_1 \cdots X_n) - H(X)| \le \epsilon] = 1$

$\Rightarrow \forall \delta, \exists n_0$, for $n \ge n_0$

$\Pr[|-\frac{1}{n}\log p(X_1, \cdots X_n) - H(X)| \le \epsilon] \in (1-\delta, 1]$

$\Pr[(X_1, \cdots X_n) \in A_\epsilon^{(n)}] \ge 1-\delta > 1-\epsilon$

$1 \ge \sum_{(x_1, x_2 \cdots x_n) \in A_\epsilon^{(n)}} p(x_1, x_2, \cdots x_n) > 1-\epsilon$ for large n

$(1-\epsilon)\, 2^{n[H(X)-\epsilon]} \le |A_\epsilon^{(n)}| \le 2^{n[H(X)+\epsilon]}$

for large n

# 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

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

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* $H(X) - \epsilon \le -\frac{1}{n}\log p(x_1, x_2, \ldots, x_n) \le H(X) + \epsilon$.
- $\Pr[A_\epsilon^{(n)}] > 1 - \epsilon$ *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.

3.

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

□

# Consequences of AEP

## Theorem 3.1.2

- *If* $(x_1, x_2, \dots, x_n) \in A_\epsilon^{(n)}$, *then* $H(X) - \epsilon \leq -\frac{1}{n}\log p(x_1, x_2, \dots, x_n) \leq H(X) + \epsilon$.
- $\Pr[A_\epsilon^{(n)}] > 1 - \epsilon$ *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.

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

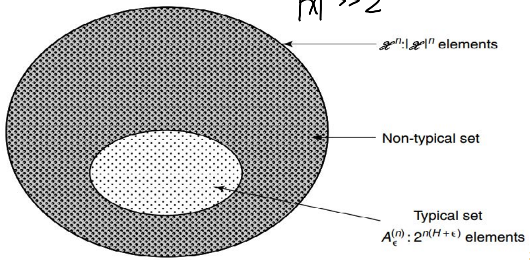
$$\begin{aligned} 1 - \epsilon &< \Pr\left[A_\epsilon^{(n)}\right] \\ &\leq \sum_{\mathbf{x} \in A_\epsilon^{(n)}} 2^{-n(H(X)-\epsilon)} \\ &= 2^{-n(H(X)-\epsilon)} \left|A_\epsilon^{(n)}\right|. \end{aligned}$$

## 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

$|\mathcal{X}|^n >> 2^{nH(X)}$



## Asymptotic Equipartition Property (AEP)

**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 $n$.*



$$X_1, X_2 \cdots X_n \begin{cases} A_\epsilon^{(n)}: \ 1 \ \underbrace{b_1 b_2 \cdots\cdots}_{\lceil n[H(X)+\epsilon]\rceil \leq n[H(X)+\epsilon]+1} \\ \overline{A_\epsilon^{(n)}}: \ 0 \ \underbrace{c_1 c_2 \cdots\cdots\cdots}_{\leq \log_2|X|^n+1} \end{cases}$$

需要的比特数.

$A_\epsilon^{(n)}: \leq n[H(X)+\epsilon]+2$

$\overline{A_\epsilon^{(n)}}: \leq n\log_2|X|+2.$

平均比特数 $\leq (1-\epsilon)\{n[H(X)+\epsilon]+2\} + \epsilon[n\log_2|X|+2]$

$\leq n[H(X)+\epsilon]+2+\epsilon[n\log_2|X|+2]$

平均描述一个变量所需的比特数 $\leq H(X) + \underbrace{\epsilon + \frac{2}{n} + \epsilon[\log_2|X| + \frac{2}{n}]}_{\epsilon'}$

# 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. $\square$

# 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. □

# Reading & Homework

Reading : 2.10 and whole Chapter 3

Homework : Problems 2.32, 3.8, 3.10