Chapter 3

Asymptotic Equipartition Property

Peng-Hua Wang

Graduate Inst. of Comm. Engineering

National Taipei University
Chapter Outline

Chap. 3 Asymptotic Equipartition Property

3.1 Asymptotic Equipartition Property Theorem

3.2 Consequences of the AEP: Data Compression

3.3 High-Probability Sets and Typical Set
3.1 Asymptotic Equipartition Property Theorem
Definition of convergence

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 \( X \)

- **In probability** if for every \( \epsilon > 0 \),

\[
\lim_{n \to \infty} \Pr \{|X_n - X| > \epsilon\} = 0
\]

or, equivalently,

\[
\lim_{n \to \infty} \Pr \{|X_n - X| < \epsilon\} = 1
\]
Definition of convergence

- **In mean square** if

\[
\lim_{n \to \infty} E[|X_n - X|^2] = 0
\]

- **With probability 1** or called **almost surely** if

\[
\Pr \left\{ \lim_{n \to \infty} X_n = X \right\} = 1
\]
Weak law of large numbers

For i.i.d. random variables $X_1, X_2, \ldots, X_n$ with common mean $m$, we have

$$\frac{1}{n} \sum_{i=1}^{n} X_i \to m \quad \text{in probability.}$$

That is, for any $\epsilon > 0$,

$$\lim_{n \to \infty} \Pr \left\{ \left| \frac{1}{n} \sum_{i=1}^{n} X_i - m \right| > \epsilon \right\} = 0$$
Theorem 3.1.1 (AEP) If $X_1, X_2, \ldots$ are i.i.d. $\sim p(x)$, then

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

Proof. Let $Z_i = -\log p(X_i)$ be i.i.d. random variables. That is, $Z_i = -\log p[X_i = x]$ if $X_i = x$, we have

$$E[Z_i] = -\sum p[X_i = x] \log p[X_i = x] = H(X_i) = H(X)$$

Now, by the weak law of large numbers,

$$\frac{1}{n} \sum_i Z_i \to H(X) \quad \text{in probability}$$

$$\Rightarrow -\frac{1}{n} \sum_i \log p(X_i) \to H(X) \quad \text{in probability}$$

$$\Rightarrow -\frac{1}{n} \log p(X_1, X_2, \ldots, X_n) \to H(X) \quad \text{in probability} \quad \square$$
Interpretation of AEP

- When $n$ is sufficient large, $p(X_1, X_2, \ldots, X_n) = 2^{-nH(X)}$ with high probability.

- For example, Let the random number $X_i$ with probability $P[X_i = 1] = p$ and $P[X_i = 0] = 1 - p = q$. If $X_1, X_2, \ldots, X_n$ are i.i.d.,

$$p(X_1, X_2, \ldots, X_n) = p^{\sum X_i} q^{n-\sum X_i}.$$ 

When $n \to \infty$,

$$p(X_1, X_2, \ldots, X_n) \to p^{np} q^{nq} = 2^{-nH}.$$ 

It means that the number of 1’s in the sequence is close to $np$, and all such sequences have roughly the same probability $2^{-nH}$. 
Interpretation of AEP

- Thus for large $n$ we can divide the sequences $X_1, X_2, \ldots, X_n$ into two types: the typical type consisting of sequences each with probability roughly $2^{-nH}$, and another type, consisting of other sequences.
Typical set

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

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

**Theorem 3.1.2**

1. If $(x_1, x_2, \ldots, x_n) \in A_{\epsilon}^{(n)}$, then

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

**Proof.** By the definition of typical set. $\square$. 
Theorems

**Theorem 3.1.2**

2. \( \Pr\{A_{\epsilon}^{(n)}\} > 1 - \epsilon \) for \( n \) sufficiently large.

**Proof.** This property follows directly from Theorem 3.1.1, since the convergence in the mean can be written as

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

Setting \( \delta = \epsilon \), we obtain the desired result. \( \square \)
Theorems

Theorem 3.1.2

3. \[ |A_{\epsilon}(n)| \leq 2^{n(H(X)+\epsilon)} \], where \(|A|\) denotes the number of elements in the set \(A\).

Proof.

\[
1 = \sum_{x \in X^n} p(x) \geq \sum_{x \in A_{\epsilon}(n)} p(x) \\
\geq \sum_{x \in A_{\epsilon}(n)} 2^{-n[H(X)+\epsilon]} = 2^{-n[H(X)+\epsilon]}|A_{\epsilon}(n)| \quad \square
\]
Theorems

Theorem 3.1.2

4. \(|A_{\epsilon}^{(n)}| \geq (1 - \epsilon)2^{n(H(X) - \epsilon)}\) for \(n\) sufficiently large.

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

\[
1 - \epsilon < \Pr\{A_{\epsilon}^{(n)}\} \leq \sum_{x \in A_{\epsilon}^{(n)}} 2^{-n[H(X) - \epsilon]} = 2^{-n[H(X) - \epsilon]}|A_{\epsilon}^{(n)}|\]

\(\Box\)
3.2 Consequences of the AEP: Data Compression
Typical set and source coding

- There are $|\mathcal{X}|^n$ elements in the whole set.
- There are $|A^{(n)}_\epsilon| \approx 2^n(H+\epsilon)$ elements in the typical set. We need $n(H+\epsilon) + 1$ bits to encode these elements, and one addition bit to indicate they are typical sequences.
- There are $|\mathcal{X}|^n - |A^{(n)}_\epsilon|$ elements in the nontypical set. We can use $n \log |\mathcal{X}| + 1$ bits to encode them, and one addition bit to indicate they are non-typical sequences.
Average length of codeword

\[ E[l(X^n)] = \sum_{x^n} p(x^n)l(x^n) \]

\[ = \sum_{x^n \in A_e^{(n)}} p(x^n)l(x^n) + \sum_{x^n \in [A_e^{(n)}]^c} p(x^n)l(x^n) \]

\[ \leq \sum_{x^n \in A_e^{(n)}} p(x^n)[n(H + \epsilon) + 2] \]

\[ + \sum_{x^n \in [A_e^{(n)}]^c} p(x^n)[n \log |\mathcal{X}| + 2] \]

\[ = \Pr\{A_e^{(n)}\}[n(H + \epsilon) + 2] + \Pr\{[A_e^{(n)}]^c\}[n \log |\mathcal{X}| + 2] \]

\[ \leq n(H + \epsilon) + \epsilon n \log |\mathcal{X}| + 2 \]

\[ = n(H + \epsilon') \]

where \( \epsilon' = \epsilon + \epsilon \log |\mathcal{X}| + \frac{2}{n} \)
Theorems

**Theorem 3.2.1** Let $X^n$ be i.i.d. $\sim p(x)$. Let $\epsilon > 0$. Then there exists a code that maps sequences $x^n$ of length $n$ into binary strings such that the mapping is one-to-one (and therefore invertible) and

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

for $n$ sufficiently large.