Elements of Information Theory

Chapter 3: Asymptotic Equipartition Property

Bilingual course
(Chinese taught course)
Information and Communication Eng. Dept.
Deng Ke

Outline

- Example
- Convergence
- Asymptotic Equipartition Property Theorem
- Strong vs. Weak Typicality
- High-probability sets and the "typical set"

Example

- For a sequence $\boldsymbol{u^L}$, whose information is

$$I(u^L)=\sum_{l=1}^{L} I(u_l)=L_1 I(s_1)+\cdots+L_K I(s_K)$$

- where $\boldsymbol{L_k}$ is the number of symbol $\boldsymbol{s_k}$
- Example: $\boldsymbol{u^L=(1, 2, 0, 3, 5, 2, 1, 2, 3)}$

$$I(u^L)=1\cdot I(0)+2\cdot I(1)+3\cdot I(2)+2\cdot I(3)+1\cdot I(5)$$

- Then

$$\begin{aligned} I(u^L) &= L\cdot\left[\frac{L_1}{L}I(s_1)+\cdots+\frac{L_K}{L}I(s_K)\right] \\ &= L\cdot\sum_k p(k)I(s_K) \qquad (L\to+\infty) \\ &= LH(U) \end{aligned}$$

- Hence, $L\to+\infty$ $\quad I(u^L)\to LH(U)$ $\quad \frac{I(u^L)}{L}\to H(U)$

Convergence of random variables

- Convergence of a sequence of numbers
- A sequence $\{a_n, n=1,2,\ldots\}$ converges to a limit a if for every $\varepsilon>0$, $\exists m$ such that $\forall n>m$, $|a_n-a|<\varepsilon$.
- A sequence of <u>random variables</u>, $\{a_n, n=1,2,\ldots\}$ converges
  - In mean square if $E(a_n-a)^2\to 0$
  - In probability if for every $\varepsilon>0$, $P\{|a_n-a|>\varepsilon\}\to 0$
  - With probability 1(also called almost surely) if $P\{\lim_{n\to\infty} a_n=a\}=1$

Law of Large Numbers

- Weak Law of Large Numbers
- Let $X_1,X_2,\ldots X_n$ be a sequence of independent and identically distributed random variables, each having finite mean $E[X_n]\to\mu$, Then, for any $\varepsilon>0$

$$\frac{X_1+X_2+\cdots X_n}{n}\to\mu \qquad \text{In probability}$$

- Strong $\frac{X_1+X_2+\cdots X_n}{n}\to\mu$ With probability 1

Asymptotic Equipartition Property

- Let $X_1,X_2,\ldots X_n$ be a sequence of i.i.d random variables with distribution $p(X)$

$$\frac{I(u^N)}{N}=-\frac{1}{N}\log p(X_1,X_2,\ldots X_N)\to H(U) \qquad \text{In probability}$$

- Typical Set

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

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

Typical Set

$$A_\varepsilon^{(n)} = \left\{ (x_1, x_2, \cdots, x_n) : \left| -\frac{1}{n} \log p(x_1, x_2 \cdots, x_n) - H(X) \right| < \varepsilon \right\}$$

- Strong and weak typical set
- $\mathcal{X}=\{a,b,c,d\}$, with probability{0.5, 0.25, 0.125, 0.125}
- Sample sequences consisting of 8 i.i.d samples
- String typical set
  - *aaaabbcd* with correct propotions
- Weak typical set
  - *aaccccc* $-\log p(x)=14=8*1.75$

Strong versus Weak Typicality

- Another example
- Bit-sequences of length n=8, $p(1)=p$, $p(0)=1-p$
- Strong typicality?
  - All sequences with about p/n 1's
- Weak typicality?
  - All sequences with probability about $2^{-n\Omega(p)}$
- What if p=0.5?

Asymptotic Equipartition Property

- All sequence in the typical set have roughly equal probabilities. Clear, a new notion of approximation is used in such a statement. We call that "exponential approximation"
- Most + least likely sequences NOT in the typical set.
- Properties of typical set
  - If $(x_1,x_2,\ldots x_n)\in A_\epsilon^{(n)}$, then

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

AEP

- 2. Pr{ $A_\epsilon^{(n)}$} > 1− ε *for n sufficiently large*
  - Proof: from the definition of AEP

$$\forall \varepsilon > 0, \delta > 0, \exists n_0 = N, \forall n > n_0$$

$$P\left(\left| -\frac{1}{n} \log p(X_1, X_2 \cdots, X_n) - H(X) \right| < \varepsilon \right) > 1 - \delta$$

$$\Rightarrow P\left(A_\varepsilon^{(n)}\right) > 1 - \varepsilon, \quad \text{if take } \delta = \varepsilon$$

  - The typical set has probability nearly 1
  - Chebyshev inequality

$$\forall \delta > 0,\ p\left[\left| -\frac{1}{n} \log p(X_1, X_2, \ldots X_n) - H(X) \right| > \delta \right] \le \frac{\sigma^2}{\delta^2},$$

$$\text{where } \sigma^2 = D\left[\frac{1}{n} \log p(X_1, X_2, \ldots X_n) \right]$$

AEP

- $\left|A_\varepsilon^{(n)}\right| \le 2^{n(H(X)+\varepsilon)}$, where $|A|$ denotes the number of elements in the set A.

$$1 = \sum_{\mathbf{x}\in\mathcal{X}^n} p(\mathbf{x}) \ge \sum_{\mathbf{x}\in A_\epsilon^{(n)}} p(\mathbf{x}) \ge \sum_{\mathbf{x}\in A_\epsilon^{(n)}} 2^{-n(H(X)+\epsilon)}$$

$$= 2^{-n(H(X)+\epsilon)} |A_\epsilon^{(n)}|,$$

- The number of elements in the typical set is nearly $2^{nH(X)}$

AEP

- $\left|A_\varepsilon^{(n)}\right| \ge (1-\varepsilon)2^{n(H(X)-\varepsilon)}$ *for n sufficiently large*

$$1 - \epsilon < \Pr\{A_\epsilon^{(n)}\} \le \sum_{\mathbf{x}\in A_\epsilon^{(n)}} 2^{-n(H(X)-\epsilon)} = 2^{-n(H(X)-\epsilon)} |A_\epsilon^{(n)}|$$

$$\sum_{\mathbf{x}\in A_\epsilon^{(n)}} p(\mathbf{x})$$

High Probability Sets

- Typical set is a fairly small set that contains most of the probability.
- It is not clear whether it is the smallest such set.
- Smallest set
- *Definition* For each n=1,2,..., let $B_\delta^{(n)} \subset X^n$ be the smallest set with
  $$P\{B_\delta^{(n)}\} > 1-\delta$$
- The smallest set is different from the typical set. must have significant intersection with and therefore must have about as many elements.

High Probability Sets

- Theorem 3.3.1 *Let* $X_1, X_2, \dots X_n$ *be* i.i.d $\sim p(x)$. *For* $\delta < 1/2$ *and any* $\delta' > 0$, *if* $P\{B_\delta^{(n)}\} > 1-\delta$, *then*

$$\frac{1}{n}\log\left|B_\delta^{(n)}\right| > H - \delta' \qquad \textit{for n sufficient large}$$

$B_\delta^{(n)}$ must have at least $2^{nH(X)}$ elements, is about the same size as the $A_\epsilon^{(n)}$

Summary

- AEP
  $$\frac{I(u^N)}{N} = -\frac{1}{N}\log p(X_1, X_2, \dots X_N) \to H(U) \quad \text{In probability}$$
- Typical Set
  $$2^{-n(H(X)+\epsilon)} \le p(x_1, x_2, \dots, x_n) \le 2^{-n(H(X)-\epsilon)}.$$
- Properties

1. If $(x_1, x_2, \dots, x_n) \in A_\epsilon^{(n)}$, then $p(x_1, x_2, \dots, x_n) = 2^{-n(H\pm\epsilon)}$.
2. $\Pr\left\{A_\epsilon^{(n)}\right\} > 1-\epsilon$ for n sufficiently large.
3. $\left|A_\epsilon^{(n)}\right| \le 2^{n(H(X)+\epsilon)}$, where $|A|$ denotes the number of elements in set A.