Lecture 3
Asymptotic Equipartition Property

Instructor: Yichen Wang

Ph.D./Professor

School of Information and Communications Engineering
Division of Electronics and Information
Xi'an Jiaotong University

Outlines

- Convergence and Law of Large Number
- Asymptotic Equipartition Property
- Typical Sequence and Typical Set
- High-Probability Sets and Typical Set

Convergence and LLN

Convergence of a sequence of numbers

A sequence $\{ a_n , n = 1, 2, 3, \ldots \}$ converges to the constant A if for every ε, there must exist an integer m such that $\forall n>m$, the following inequality holds:

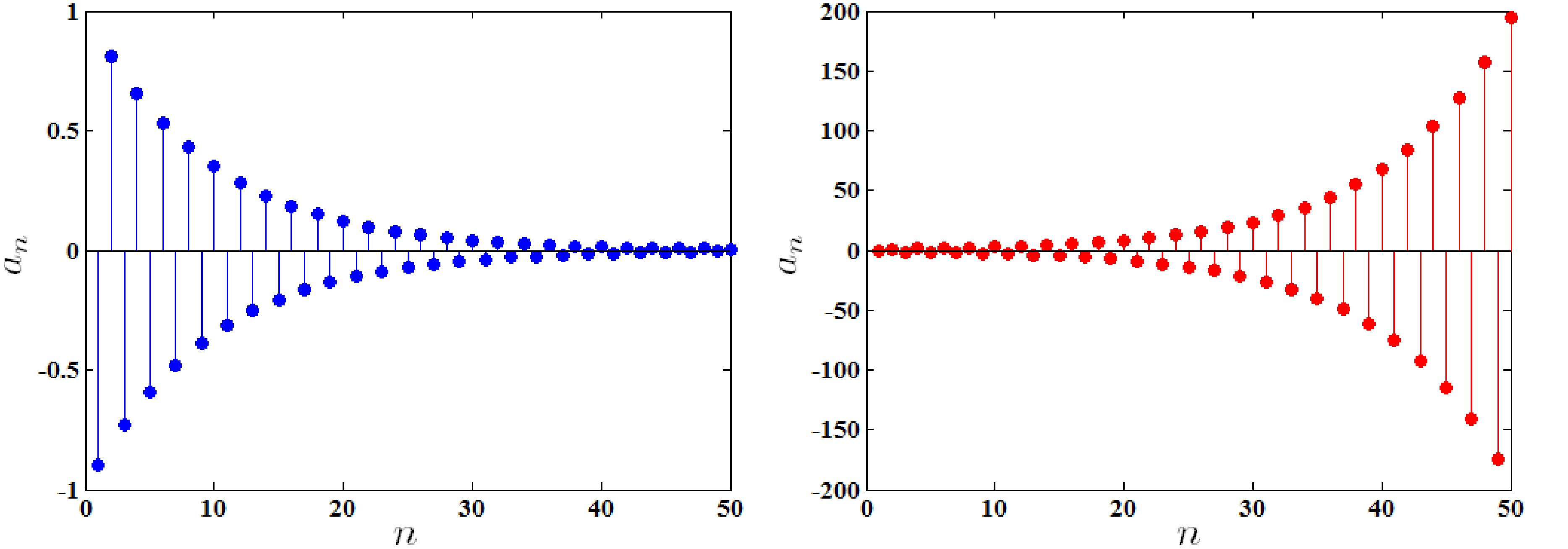
$$|a_n - A| < \varepsilon$$

Question:

If the sequence $\{ a_n , n = 1, 2, 3, \ldots \}$ converges to the constant A, what does it imply?

Convergence and LLN

$$a_n = \left(-\frac{9}{10}\right)^n, \; n = 1, 2, 3, \cdots \qquad a_n = \left(-\frac{10}{9}\right)^n, \; n = 1, 2, 3, \cdots$$



Can we extend the definition of the convergence for the sequence of number to the random variable case?

Convergence and LLN

Convergence of a sequence of random variables

- *It can also be called as "Stochastic Convergence".*
- *It is an critically important concept in probabilities as well as other applications.*
- *Formalizing the idea that a sequence of essentially random or unpredictable events can sometimes be expected to settle into a pattern.*
- *There are several approaches to characterize the convergence of the sequence of random variables (stochastic convergence).*

Convergence and LLN

A sequence of random variable { X_n , n = 1, 2, 3, …} can be viewed as converging to the random variable X

1) Convergence in distribution

$$\lim_{n\to\infty} F_n(x) = F(x), \quad \text{where } F_n \text{ and } F \text{ are CDFs of } X_n \text{ and } X$$

2) Convergence in probability

$$\lim_{n\to\infty} \Pr\{|X_n - X| \geq \varepsilon\} = 0$$

3) Almost sure convergence (with probability 1)

$$\Pr\left\{\lim_{n\to\infty} X_n = X\right\} = 1$$

4) Convergence in mean

$$\lim_{n\to\infty} \mathbb{E}\left\{|X_n - X|^r\right\} = 0, \quad \text{where } r \text{ is the real number and } r \geq 1$$

$$\text{when } r = 2 \Longrightarrow \text{Convergence in mean square}$$

Convergence and LLN

Law of Large Numbers (LLN)

- ***It describes the result of performing the same experiment a large number of times.***

- ***It was first stated by Italian mathematician Gerolamo Cardano.***

- ***A special form of LLN (binary random variable) was proved by Jacob Bernoulli ------ "Bernoulli's Theorem"***

- ***It was further described by S. D. Poisson and named as LLN.***

- ***Weak LLN and Strong LLN***

Convergence and LLN

Law of Large Numbers (LLN)

Let X_1, X_2, ……, X_n be a sequence of independent and identically distributed (i.i.d.) random variables with expected (mean) value:

$$\mathbb{E}\{X_1\} = \mathbb{E}\{X_2\} = \cdots = \mathbb{E}\{X_n\} = \mu$$

Then, the LLN tells us that the sample average

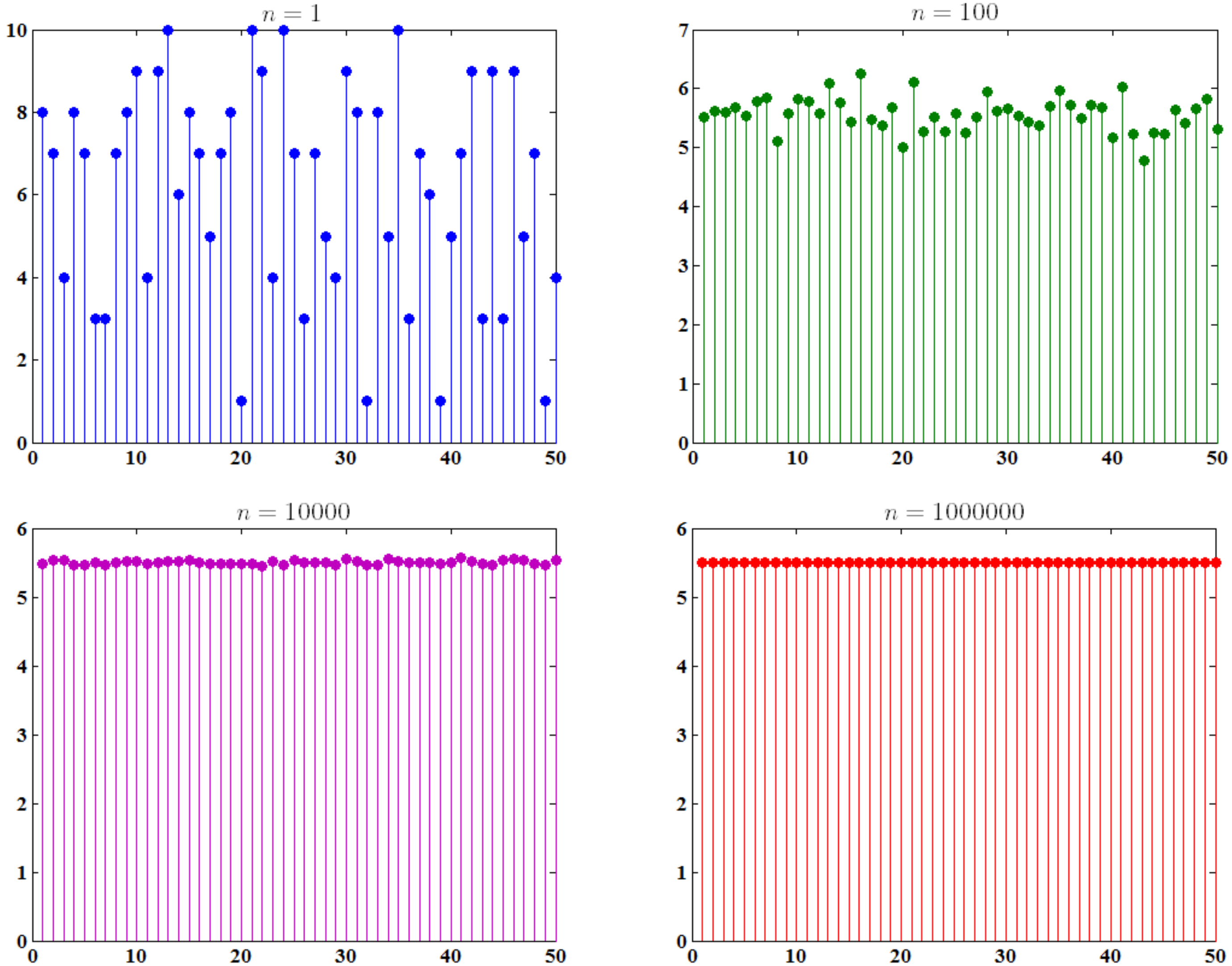
$$\overline{X}_n = \frac{1}{n}\left(X_1 + X_2 + \cdots + X_n\right)$$

converges to the expected value, i.e.,

$$\overline{X}_n \longrightarrow \mu \quad \text{for} \quad n \to \infty$$

Convergence and LLN

X_n is uniformly distributed on {1, 2, …, 10}. $\overline{X}_n = \dfrac{X_1 + \cdots + X_n}{n}$



Convergence and LLN

➢ ***Weak Law***

The sample value of the sequence of i.i.d. random variables X_1, X_2, ……, X_n converges in probability towards the expected value. That is to say for any positive number ε, we have

$$\lim_{n\to\infty} \Pr\left\{ \left|\overline{X}_n - \mu\right| > \varepsilon \right\} = 0$$

➢ ***Strong Law***

The sample value of the sequence of i.i.d. random variables X_1, X_2, ……, X_n converges almost surely to the expected value, i.e.,

$$\Pr\left\{ \lim_{n\to\infty} \overline{X}_n = \mu \right\} = 1$$

Outlines

- Convergence and Law of Large Number
- Asymptotic Equipartition Property
- Typical Sequence and Typical Set
- High-Probability Sets and Typical Set

Asymptotic Equipartition Property

Example

We consider a sequence

$$\boldsymbol{u}_L = (u_1, u_2, \cdots, u_L),\ \ u_l \in \{s_1, \cdots, s_K\}$$

Then, the *self-information* of this sequence is

$$I\left(\boldsymbol{u}_L\right) = \sum_{l=1}^{L} I(u_l) = L_1 I(s_1) + L_2 I(s_2) + \cdots + L_K I(s_K),$$

where L_k denotes the number of symbol s_k in the sequence.

For example, we consider the following sequence

$$\boldsymbol{u}_L = (1, 2, 0, 3, 5, 2, 1, 2, 3)$$

Then, we have

$$I\left(\boldsymbol{u}_L\right) = 1 \cdot I(0) + 2 \cdot I(1) + 3 \cdot I(2) + 2 \cdot I(3) + 1 \cdot I(5)$$

Asymptotic Equipartition Property

$$I(\boldsymbol{u}_L) = \sum_{l=1}^{L} I(u_l) = L_1 I(s_1) + L_2 I(s_2) + \cdots + L_K I(s_K)$$

$$\frac{1}{L} I(\boldsymbol{u}_L) = \boxed{\frac{1}{L}\sum_{l=1}^{L} I(u_l)} = \frac{L_1}{L} I(s_1) + \frac{L_2}{L} I(s_2) + \cdots + \frac{L_K}{L} I(s_K)$$

$L \to \infty$

Converging to the expected value of $I(u_l)$

Questions

1. ***What is the expected value of the self-information?***
2. ***What can we obtain from this example?***

Asymptotic Equipartition Property

Theorem (Asymptotic Equipartition Property, AEP)

Let $X_1, X_2, \ldots, X_n$ be the sequence of i.i.d. random variables with distribution p(x), then we have

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

i.e.,

$$\lim_{n\to\infty} \Pr\left\{ \left| -\frac{1}{n}\log p(X_1, X_2, \cdots, X_n) - H(X) \right| > \varepsilon \right\} = 0, \quad \forall\varepsilon > 0.$$

It is derived by weak LLN. $\longrightarrow$ ***Weak AEP***

Asymptotic Equipartition Property

Strong LLN $\longrightarrow$ Strong AEP

Strong AEP

Let $X_1, X_2, \ldots, X_n$ be the sequence of i.i.d. random variables with distribution $p(x)$, then we have

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

i.e.,

$$\Pr\left\{\lim_{n\to\infty} -\frac{1}{n}\log p(X_1, X_2, \cdots, X_n) = H(X)\right\} = 1.$$

Questions:

1. ***What is the relationship between LLN and AEP?***
2. ***Why is it named as AEP?***

Outlines

- Convergence and Law of Large Number
- Asymptotic Equipartition Property
- Typical Sequence and Typical Set
- High-Probability Sets and Typical Set

Typical Sequence and Typical Set

We consider a sequence of i.i.d. random variables

$$\boldsymbol{X}_L = (X_1, X_2, \cdots, X_L),\ \ X_i \in \{s_1, \cdots, s_K\}$$

with distribution *p*(*x*).

Then, the probability of event (x_1, x_2, …, x_L) can be written as

The number of the corresponding symbol appeared in the sequence

$$\Pr\Big\{X_1 = x_1, X_2 = x_2, \cdots, X_L = x_L\Big\}$$

$$= \Big(\Pr\Big\{X = s_1\Big\}\Big)^{L_1}\Big(\Pr\Big\{X = s_2\Big\}\Big)^{L_2}\cdots\Big(\Pr\Big\{X = s_K\Big\}\Big)^{L_K}$$

Different sequences usually have distinct probabilities.

Typical Sequence and Typical Set

Different sequences usually have distinct probabilities.

$$\Pr\left\{X_1 = x_1, X_2 = x_2, \cdots, X_L = x_L\right\}$$

$$= \left(\Pr\left\{X = s_1\right\}\right)^{L_1} \left(\Pr\left\{X = s_2\right\}\right)^{L_2} \cdots \left(\Pr\left\{X = s_K\right\}\right)^{L_K}$$

For example, we consider the random variable X $\in$ {0,1} with P(0) = 0.3 and P(1) = 0.7. Then, for the following two sequences, we have

$$\Pr\left\{\boldsymbol{x}_9^1 = (1,1,1,1,1,1,0,0,0)\right\} = 0.0032$$

$$\Pr\left\{\boldsymbol{x}_9^2 = (0,0,0,0,0,0,1,1,1)\right\} = 0.00025$$

What will happen if L goes to infinite?

Typical Sequence and Typical Set

Weak Asymptotic Equipartition Property

$$\lim_{L\to\infty} \Pr\left\{ \left| -\frac{1}{L}\log P(X_1, X_2, \cdots, X_L) - H(X) \right| \le \varepsilon \right\} = 1,\ \forall\varepsilon > 0$$

$$2^{-L(H(X)+\varepsilon)} \le P(X_1, X_2, \cdots, X_L) \le 2^{-L(H(X)-\varepsilon)},\ \forall\varepsilon > 0 \text{ and } L \to \infty$$

When L➔ ∞, different sequences have approximately the same probabilities.

Approximately equal partition on probability

Typical Sequence and Typical Set

We consider a sequence of i.i.d. random variables

$$\boldsymbol{X}_L = (X_1, X_2, \cdots, X_L),\ X_i \in \{s_1, \cdots, s_K\}$$

with distribution *p*(*x*).

For different realizations, the numbers of symbols s_1, … s_K are different.

For example, we consider the following two sequences

$$\boldsymbol{x}_9^1 = (1, 1, 1, 1, 1, 1, 0, 0, 0)$$

$$\boldsymbol{x}_9^2 = (0, 0, 0, 0, 0, 0, 1, 1, 1)$$

What will happen if L goes to infinite?

Typical Sequence and Typical Set

Strong Asymptotic Equipartition Property

$\downarrow$

Strong Law of Large Numbers

$\downarrow$

$$\Pr\left\{ \lim_{L\to\infty} \frac{1}{L} \sum_{i=1}^{L} X_i = \mathbb{E}\{X\} \right\} = 1$$

$\downarrow$

$$\forall\ \mathbf{x}_L = (x_1, \cdots, x_L) \Longrightarrow \frac{1}{L} \sum_{i=1}^{L} x_i = \mathbb{E}\{X\},\ \text{if } L \to \infty$$

$$\sum_{k=1}^{K} \left[\frac{L_i}{L} - P(s_k) \right] s_k = 0,\ \text{if } L \to \infty$$

The empirical distribution in sequence x_L should be "close" to the prior probability distribution.

Typical Sequence and Typical Set

We consider a sequence of i.i.d. random variables

$$\boldsymbol{X}_L = (X_1, X_2, \cdots, X_L),\ \ X_i \in \{s_1, \cdots, s_K\}$$

with distribution *p*(*x*).

Then, the probability of event (x_1, x_2, …, x_L) can be written as

$$\Pr\left\{X_1 = x_1, X_2 = x_2, \cdots, X_L = x_L\right\}$$

$$= \left(\Pr\left\{X = s_1\right\}\right)^{L_1}\left(\Pr\left\{X = s_2\right\}\right)^{L_2}\cdots\left(\Pr\left\{X = s_K\right\}\right)^{L_K}$$

Different sequences usually have distinct probabilities.

Typical Sequence and Typical Set

Different sequences usually have distinct probabilities.

$$\Pr\left\{X_1 = x_1, X_2 = x_2, \cdots, X_L = x_L\right\}$$

$$= \left(\Pr\left\{X = s_1\right\}\right)^{L_1} \left(\Pr\left\{X = s_2\right\}\right)^{L_2} \cdots \left(\Pr\left\{X = s_K\right\}\right)^{L_K}$$

Strong Asymptotic Equipartition Property

$$\Downarrow$$

Empirical distribution = Prior probability distribution

$$\Downarrow$$

Constant and equal to $2^{-LH(X)}$ when L goes to infinity.

Typical Sequence and Typical Set

Definition (Weakly Typical Sequence)

Let X ~ p(x). Then, the sequence

$$\left(x_1, x_2, \cdots, x_n\right) \in \mathcal{X}^n$$

is weakly typical sequence if

$$\left|-\frac{1}{n}\log p(x_1, x_2, \cdots, x_n) - H(X)\right| \leq \varepsilon.$$

Definition (Strongly Typical Sequence)

Let X ~ p(x). Then, the sequence

$$\left(x_1, x_2, \cdots, x_n\right) \in \mathcal{X}^n$$

is strongly typical sequence if the empirical distribution of each symbol is "close" to its prior probability distribution.

Typical Sequence and Typical Set

Definition (Weakly Typical Set)

The weakly typical set $A_\epsilon^{(n)}$ with respect to p(x) is the set of sequences

$$\left(x_1, x_2, \cdots, x_n\right) \in \mathcal{X}^n$$

with the property

$$\left| -\frac{1}{n}\log p(x_1, x_2, \cdots, x_n) - H(X) \right| \leq \varepsilon.$$

Definition (Strongly Typical Set)

The strongly typical set $\mathcal{T}_\epsilon^{(n)}$ with respect to p(x) is the set of sequences

$$\left(x_1, x_2, \cdots, x_n\right) \in \mathcal{X}^n$$

with the property

$$\forall\ a \in \mathcal{X}, \quad \left| \frac{1}{n} N\left(a; x^n\right) - p(a) \right| \leq \epsilon p(a),$$

where $N(a;x^n)$ is the number of occurrences of symbol a in sequence.

Typical Sequence and Typical Set

- ***Strong typical set v.s. Weak typical set***
  - ***Strong typical set – Strong typical sequence***
  - ***Weak typical set – Weak typical sequence***

Example

- ✓ ***X = {a,b,c,d} with probability distribution {0.5, 0.25, 0.125, 0.125}***
- ✓ ***Sample sequences consisting of 8 i.i.d. samples***
- ✓ ***Strong typical sequence***
  ---- Symbols with correct proportions, e.g., aaaabbcd
- ✓ ***Weak typical sequence***
  ---- Self-information per symbol equals to the entropy, e.g., aabbbbbb

Typical Sequence and Typical Set

Example

- ✓ *Bit-sequences of length n*
- ✓ *The probability distribution is $P(1) = p$ and $P(0) = 1-p$*
- ✓ *Strong typicality?*
- ✓ *Weak typicality?*
- ✓ *What will happen when $p=0.5$?*

Some Discussions

1. *All sequence in the typical set have roughly equal probabilities.*
2. *Most or least likely sequences may NOT be in the typical set.*

Typical Sequence and Typical Set

Properties of (Weakly) Typical Set

1. If $(x_1, x_2, \cdots, x_n) \in A_\epsilon^{(n)}$ ***, then we have***

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

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 the set A.***

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

Typical Sequence and Typical Set

1. ***If*** $(x_1, x_2, \cdots, x_n) \in A_\epsilon^{(n)}$***, then we can obtain from the definition of weak typical set that***

$$\left| -\frac{1}{n}\log p(x_1, x_2, \cdots, x_n) - H(X) \right| \le \varepsilon$$

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

2. ***Based on weak AEP, we have***

$$\lim_{n\to\infty} \Pr\left\{ \left| -\frac{1}{n}\log p(X_1, X_2, \cdots, X_n) - H(X) \right| > \varepsilon \right\} = 0, \quad \forall \varepsilon > 0$$

For any given $\delta > 0$***, there must exist*** n_0 ***such that when*** $n > n_0$***, we have***

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

Let $\delta = \varepsilon$***, we have*** $\Pr\left\{ A_\epsilon^{(n)} \right\} > 1 - \epsilon$.

Typical Sequence and Typical Set

3. Proof for property (3)

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

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

$$1=\sum_{\boldsymbol{x}\in\mathcal{X}^n} p(\boldsymbol{x}) \ge \sum_{\boldsymbol{x}\in A_\varepsilon^{(n)}} p(\boldsymbol{x}) \ge \sum_{\boldsymbol{x}\in A_\varepsilon^{(n)}} 2^{-n[H(X)+\epsilon]} = 2^{-n[H(X)+\epsilon]}\left|A_\varepsilon^{(n)}\right|$$

4. Proof for property (4)

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

Outlines

- Convergence and Law of Large Number
- Asymptotic Equipartition Property
- Typical Sequence and Typical Set
- High-Probability Sets and Typical Set

High- Probability Sets and Typical Set

- **Typical set is a fairly small set that contains most of the probability.**
- **It is not clear whether it is the smallest such set.**

Definition (Smallest Set)

For each n=1,2,…, let $B_\delta^{(n)} \subset \mathcal{X}^n$ ***be the smallest set with***

$$\Pr\left\{B_\delta^{(n)}\right\} \geq 1-\delta$$

- ***The smallest set is different from the typical set.***
- ***The smallest set must have significant intersection with the typical set and therefore must have about as many elements.***

High- Probability Sets and Typical Set

Theorem

Let X_1, X_2, … , X_n be i.i.d. random variables with probability distribution p(x). For $\delta < 1/2$ and any $\delta' > 0$, if

$$\Pr\left\{B_\delta^{(n)}\right\} \geq 1-\delta$$

is satisfied, that we have

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

for n sufficiently large.

The theorem tells us $B_\delta^{(n)}$ must have at least 2^{nH} elements and is about the same size as the $A_\epsilon^{(n)}$.

Summary

- **AEP**

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

- **Typical Set**

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

- **Properties**

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

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

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

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