Chapter 5
Weak Typicality

Department of Information Engineering
The Chinese University of Hong Kong

The Notion of Typical Sequences

- Toss a fair coin n times.
- If the outcome is "head" approximately half of the time, the sequence of outcomes is "normal", or "typical".
- How to measure the typicality of a sequence w.r.t. to a generic distribution of an i.i.d. process?
- Two common such measures in information theory: weak typicality and strong typicality.
- The main theorems are weak and strong *Asymptotic Equipartition Properties* (AEP), which are consequences of WLLN.

5.1 The Weak AEP

Setup

- $\{X_k, k \geq 1\}$, X_k i.i.d. $\sim p(x)$.
- X denotes generic r.v. with entropy $H(X) < \infty$.
- $\mathbf{X} = (X_1, X_2, \cdots, X_n)$. Then

$$p(\mathbf{X}) = p(X_1)p(X_2)\cdots p(X_n).$$

- $\mathcal{X}$ may be countably infinite.

Theorem 5.1 (Weak AEP I)

$$-\frac{1}{n}\log p(\mathbf{X}) \to H(X)$$

in probability as $n \to \infty$, i.e., for any $\epsilon > 0$, for n sufficiently large,

$$\Pr\left\{\left|-\frac{1}{n}\log p(\mathbf{X}) - H(X)\right| \leq \epsilon\right\} > 1 - \epsilon.$$

Note: $X_n \to X$ in probability means that

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

for all $\epsilon > 0$.

Definition 5.2 The weakly typical set $W^n_{[X]\epsilon}$ with respect to $p(x)$ is the set of sequences $\mathbf{x} = (x_1, x_2, \cdots, x_n) \in \mathcal{X}^n$ such that

$$\left| -\frac{1}{n} \log p(\mathbf{x}) - H(X) \right| \leq \epsilon,$$

or equivalently,

$$H(X) - \epsilon \leq -\frac{1}{n} \log p(\mathbf{x}) \leq H(X) + \epsilon,$$

where ϵ is an arbitrarily small positive real number. The sequences in $W^n_{[X]\epsilon}$ are called weakly ϵ-typical sequences.

Empirical Entropy

- $$-\frac{1}{n}\log p(\mathbf{x}) = -\frac{1}{n}\sum_{k=1}^{n}\log p(x_k)$$

  is called the *empirical entropy* of the sequence $\mathbf{x}$.

- The empirical entropy of a weakly typical sequence is close to the true entropy $H(X)$.

Theorem 5.2 (Weak AEP II) The following hold for any $\epsilon > 0$:

1) If $\mathbf{x} \in W^n_{[X]\epsilon}$, then

$$2^{-n(H(X)+\epsilon)} \leq p(\mathbf{x}) \leq 2^{-n(H(X)-\epsilon)}.$$

2) For n sufficiently large,

$$\Pr\{\mathbf{X} \in W^n_{[X]\epsilon}\} > 1 - \epsilon.$$

3) For n sufficiently large,

$$(1-\epsilon)2^{n(H(X)-\epsilon)} \leq |W^n_{[X]\epsilon}| \leq 2^{n(H(X)+\epsilon)}.$$

WAEP says that for large n,

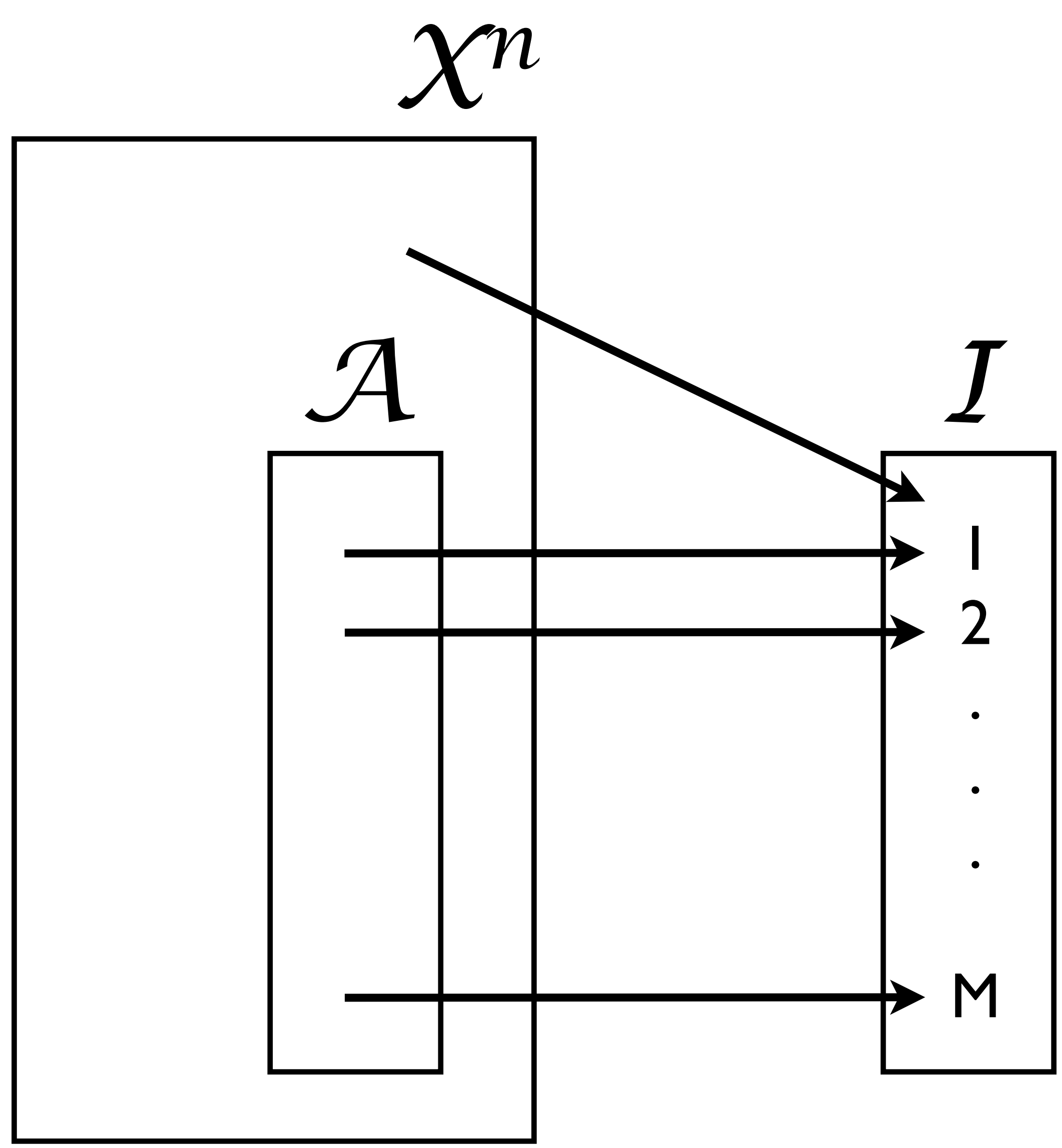
- the probability of occurrence of the sequence drawn is close to $2^{-nH(X)}$ with very high probability;
- the total number of weakly typical sequences is approximately equal to $2^{nH(X)}$.

WAEP DOES NOT say that

- most of the sequences in $\mathcal{X}^n$ are weakly typical;
- the most likely sequence is weakly typical.

When n is large, one can almost think of the sequence $\mathbf{X}$ as being obtained by choosing a sequence from the weakly typical set according to the uniform distribution.

5.2 The Source Coding Theorem



A block code: $\mathcal{X}^n \to \mathcal{I}$

- $\mathcal{I} = \{1, 2, \cdots, M\}$
- blocklength $= n$
- coding rate $= n^{-1} \log M$
- $P_e = \Pr\{\mathbf{X} \notin \mathcal{A}\}$.

Direct Part & Converse

- Direct part: For arbitrarily small P_e, there exists a block code whose coding rate is arbitrarily close to $H(X)$ when n is sufficiently large.

- Converse: For any block code with block length n and coding rate less than $H(X) - \zeta$, where $\zeta > 0$ does not change with n, then $P_e \to 1$ as $n \to \infty$.

Direct Part

- Fix $\epsilon > 0$ and take $\mathcal{A} = W^n_{[X]\epsilon}$ and hence $M = |\mathcal{A}|$.
- For sufficiently large n, by WAEP,

$$(1-\epsilon)2^{n(H(X)-\epsilon)} \leq M = |\mathcal{A}| = |W^n_{[X]\epsilon}| \leq 2^{n(H(X)+\epsilon)}.$$

- Coding rate satisfies

$$\frac{1}{n}\log(1-\epsilon) + H(X) - \epsilon \leq \frac{1}{n}\log M \leq H(X) + \epsilon.$$

- By WAEP,

$$P_e = \Pr\{\mathbf{X} \notin \mathcal{A}\} = \Pr\{\mathbf{X} \notin W^n_{[X]\epsilon}\} < \epsilon.$$

- Letting $\epsilon \to 0$, the coding rate tends to $H(X)$, while P_e tends to 0.

Converse

- Consider any block code with rate less than $H(X) - \zeta$, where $\zeta > 0$ does not change with n. Then total number of codewords $\leq 2^{n(H(X)-\zeta)}$.
- Use some indices to cover $\mathbf{x} \in W^n_{[X]\epsilon}$, and others to cover $\mathbf{x} \notin W^n_{[X]\epsilon}$.
- Total probability of typical sequences covered is upper bounded by

$$2^{n(H(X)-\zeta)} 2^{-n(H(X)-\epsilon)} = 2^{-n(\zeta-\epsilon)}.$$

- Total probability covered is upper bounded by

$$2^{-n(\zeta-\epsilon)} + \Pr\{\mathbf{X} \notin W^n_{[X]\epsilon}\} < 2^{-n(\zeta-\epsilon)} + \epsilon.$$

- Then $P_e > 1 - (2^{-n(\zeta-\epsilon)} + \epsilon)$ holds for any $\epsilon > 0$ and n sufficiently large.
- Take $\epsilon < \zeta$. Then $P_e > 1 - 2\epsilon$ for n sufficiently large.
- Finally, let $\epsilon \to 0$.