Chapter 2

General Data Compression Theorems

Po-Ning Chen

Institute of Communications Engineering

National Chiao-Tung University

Hsin Chu, Taiwan 30010

Motivations

- We already know that the entropy rate

$$\lim_{n\to\infty} \frac{1}{n} H(X^n)$$

is the minimum data compression rate for arbitrarily small data compression error for block coding of the stationary ergodic source.

- We also mentioned that for a more complicated situation where the source becomes non-stationary, the quantity $\lim_{n\to\infty}(1/n)H(X^n)$ may not exist, and can no longer be used to characterize the source compression.

- This results in the need to establish a new entropy measure which appropriately characterizes the operational limits of arbitrary stochastic systems, which was done in the previous chapter.

Fixed-length codes for arbitrary sources

- Here, we have made an implicit assumption in the following derivation, which is the source alphabet $\mathcal{X}$ is *finite*.

Definition 2.1 (cf. Definition 3.2 and its associated footnote in [2])
An (n, M) block code for data compression is a set

$$\mathscr{C}_n := \{\boldsymbol{c}_1, \boldsymbol{c}_2, \ldots, \boldsymbol{c}_M\}$$

consisting of M sourcewords of block length n (and a binary-indexing codeword for each sourceword $\boldsymbol{c}_i$).

Fixed-length codes for arbitrary sources

Definition 2.2 Fix $\varepsilon \in [0,1]$. R is an ε-achievable data compression rate for a source $\boldsymbol{X}$ if there exists a sequence of block data compression codes $\{\mathscr{C}_n = (n, M_n)\}_{n=1}^{\infty}$ with

$$\limsup_{n\to\infty} \frac{1}{n} \log M_n \le R,$$

and

$$\limsup_{n\to\infty} P_e(\mathscr{C}_n) \le \varepsilon,$$

where $P_e(\mathscr{C}_n) := \Pr\left(X^n \notin \mathscr{C}_n\right)$ is the probability of decoding error.

The infimum of all ε-achievable data compression rate for $\boldsymbol{X}$ is denoted by $T_\varepsilon(\boldsymbol{X})$.

- Note that in conventional source coding theorem, one wants to find the minimum rate with arbitrary small error. This rate is exactly $\lim_{\varepsilon\downarrow 0} T_\varepsilon(\boldsymbol{X})$.
- As expected, for DMS, $\lim_{\varepsilon\downarrow 0} T_\varepsilon(\boldsymbol{X}) = H(X)$. Actually, for DMS, $T_\varepsilon(\boldsymbol{X}) = H(X)$ for any $\varepsilon \in [0,1)$.

Fixed-length codes for arbitrary sources

Lemma 2.3 Fix a positive integer n. There exists an (n, M_n) source block code $\mathscr{C}_n$ for P_{X^n} such that its error probability satisfies

$$P_e(\mathscr{C}_n) \le \Pr\left[\frac{1}{n} h_{X^n}(X^n) > \frac{1}{n}\log M_n\right].$$

Proof: Observe that

$$\begin{aligned} 1 &\ge \sum_{\{x^n\in\mathcal{X}^n\ :\ (1/n)h_{X^n}(x^n)\le(1/n)\log M_n\}} P_{X^n}(x^n) \\ &\ge \sum_{\{x^n\in\mathcal{X}^n\ :\ (1/n)h_{X^n}(x^n)\le(1/n)\log M_n\}} \frac{1}{M_n} \\ &\ge \left|\left\{x^n \in \mathcal{X}^n\ :\ \frac{1}{n}h_{X^n}(x^n) \le \frac{1}{n}\log M_n\right\}\right| \frac{1}{M_n}. \end{aligned}$$

Therefore, $|\{x^n \in \mathcal{X}^n\ :\ (1/n)h_{X^n}(x^n) \le (1/n)\log M_n\}| \le M_n$. We can then choose a code

$$\mathscr{C}_n \supset \left\{x^n \in \mathcal{X}^n\ :\ \frac{1}{n}h_{X^n}(x^n) \le \frac{1}{n}\log M_n\right\}$$

with $|\mathscr{C}_n| = M_n$ and

$$P_e(\mathscr{C}_n) = 1 - P_{X^n}\{\mathscr{C}_n\} \le \Pr\left[\frac{1}{n}h_{X^n}(X^n) > \frac{1}{n}\log M_n\right].$$

□

Fixed-length codes for arbitrary sources

Lemma 2.4 Every (n, M_n) source block code $\mathcal{C}_n$ for P_{X^n} satisfies

$$P_e(\mathcal{C}_n) \geq \Pr\left[\frac{1}{n} h_{X^n}(X^n) > \frac{1}{n}\log M_n + \gamma\right] - \exp\{-n\gamma\},$$

for every $\gamma > 0$.

Proof: It suffices to prove that

$$1 - P_e(\mathcal{C}_n) = \Pr\{X^n \in \mathcal{C}_n\} \leq \Pr\left[\frac{1}{n} h_{X^n}(X^n) \leq \frac{1}{n}\log M_n + \gamma\right] + \exp\{-n\gamma\}.$$

Clearly,

$$\begin{aligned}\Pr\{X^n \in \mathcal{C}_n\} \;&=\; \Pr\left\{X^n \in \mathcal{C}_n \text{ and } \frac{1}{n} h_{X^n}(X^n) \leq \frac{1}{n}\log M_n + \gamma\right\} \\ &\quad + \Pr\left\{X^n \in \mathcal{C}_n \text{ and } \frac{1}{n} h_{X^n}(X^n) > \frac{1}{n}\log M_n + \gamma\right\}\end{aligned}$$

Fixed-length codes for arbitrary sources

$$
\begin{aligned}
&\le \Pr\left\{\frac{1}{n}h_{X^n}(X^n) \le \frac{1}{n}\log M_n + \gamma\right\} \\
&\quad + \Pr\left\{X^n \in \mathcal{C}_n \text{ and } \frac{1}{n}h_{X^n}(X^n) > \frac{1}{n}\log M_n + \gamma\right\} \\
&= \Pr\left\{\frac{1}{n}h_{X^n}(X^n) \le \frac{1}{n}\log M_n + \gamma\right\} \\
&\quad + \sum_{x^n \in \mathcal{C}_n} P_{X^n}(x^n) \cdot \mathbf{1}\left\{\frac{1}{n}h_{X^n}(x^n) > \frac{1}{n}\log M_n + \gamma\right\} \\
&= \Pr\left\{\frac{1}{n}h_{X^n}(X^n) \le \frac{1}{n}\log M_n + \gamma\right\} \\
&\quad + \sum_{x^n \in \mathcal{C}_n} P_{X^n}(x^n) \cdot \mathbf{1}\left\{P_{X^n}(x^n) < \frac{1}{M_n}\exp\{-n\gamma\}\right\} \\
&< \Pr\left\{\frac{1}{n}h_{X^n}(X^n) \le \frac{1}{n}\log M_n + \gamma\right\} + |\mathcal{C}_n|\frac{1}{M_n}\exp\{-n\gamma\} \\
&= \Pr\left\{\frac{1}{n}h_{X^n}(X^n) \le \frac{1}{n}\log M_n + \gamma\right\} + \exp\{-n\gamma\}.
\end{aligned}
$$

□

Fixed-length codes for arbitrary sources

We now apply Lemmas 2.3 and 2.4 to prove a *general* source coding theorems for block codes.

Theorem 2.5 (general source coding theorem) For any source $\boldsymbol{X}$,

$$T_\varepsilon(\boldsymbol{X}) = \begin{cases} \lim_{\delta\uparrow(1-\varepsilon)} \bar{H}_\delta(\boldsymbol{X}), & \text{for } \varepsilon \in [0,1); \\ 0, & \text{for } \varepsilon = 1. \end{cases}$$

Proof: The case of $\varepsilon = 1$ follows directly from its definition; hence, the proof only focus on the case of $\varepsilon \in [0,1)$.

1. . *Forward part (achievability):* $T_\varepsilon(\boldsymbol{X}) \le \lim_{\delta\uparrow(1-\varepsilon)} \bar{H}_\delta(\boldsymbol{X})$

   We need to prove the existence of a sequence of block codes $\{\mathscr{C}_n = (n, M_n)\}_{n\ge1}$ such that for every $\gamma > 0$,

$$\limsup_{n\to\infty} \frac{1}{n}\log M_n \le \lim_{\delta\uparrow(1-\varepsilon)} \bar{H}_\delta(\boldsymbol{X}) + \gamma \quad \text{and} \quad \limsup_{n\to\infty} P_e(\mathscr{C}_n) \le \varepsilon.$$

   Lemma 2.3 ensures the existence (for any $\gamma > 0$) of a source block code $\mathscr{C}_n = (n, M_n = \lceil \exp\{n(\lim_{\delta\uparrow(1-\varepsilon)} \bar{H}_\delta(\boldsymbol{X}) + \gamma)\}\rceil)$ with error probability

$$\begin{aligned} P_e(\mathscr{C}_n) &\le \Pr\left\{\frac{1}{n}h_{X^n}(X^n) > \frac{1}{n}\log M_n\right\} \\ &\le \Pr\left\{\frac{1}{n}h_{X^n}(X^n) > \lim_{\delta\uparrow(1-\varepsilon)} \bar{H}_\delta(\boldsymbol{X}) + \gamma\right\}. \end{aligned}$$

Therefore,

$$\begin{aligned}\limsup_{n\to\infty} P_e(\mathcal{C}_n) &\leq \limsup_{n\to\infty} \Pr\left\{\frac{1}{n} h_{X^n}(X^n) > \lim_{\delta\uparrow(1-\varepsilon)} \bar{H}_\delta(\boldsymbol{X}) + \gamma\right\} \\ &= 1 - \liminf_{n\to\infty} \Pr\left\{\frac{1}{n} h_{X^n}(X^n) \leq \lim_{\delta\uparrow(1-\varepsilon)} \bar{H}_\delta(\boldsymbol{X}) + \gamma\right\} \\ &\leq 1 - (1-\varepsilon) = \varepsilon,\end{aligned}$$

where the last inequality follows from

$$\lim_{\delta\uparrow(1-\varepsilon)} \bar{H}_\delta(\boldsymbol{X}) = \sup\left\{\theta : \liminf_{n\to\infty} \Pr\left[\frac{1}{n} h_{X^n}(X^n) \leq \theta\right] < 1-\varepsilon\right\}. \quad (2.1.1)$$

2. *Converse part:* $T_\varepsilon(\boldsymbol{X}) \geq \lim_{\delta\uparrow(1-\varepsilon)} \bar{H}_\delta(\boldsymbol{X})$

Assume without loss of generality that $\lim_{\delta\uparrow(1-\varepsilon)} \bar{H}_\delta(\boldsymbol{X}) > 0$. We will prove the converse by contradiction. Suppose that $T_\varepsilon(\boldsymbol{X}) < \lim_{\delta\uparrow(1-\varepsilon)} \bar{H}_\delta(\boldsymbol{X})$. Then $(\exists\ \gamma > 0)$ $T_\varepsilon(\boldsymbol{X}) < \lim_{\delta\uparrow(1-\varepsilon)} \bar{H}_\delta(\boldsymbol{X}) - 4\gamma$. By definition of $T_\varepsilon(\boldsymbol{X})$, there exists a sequence of codes $\mathcal{C}_n = (n, M_n)$ such that

$$\limsup_{n\to\infty} \frac{1}{n} \log M_n \leq \left(\lim_{\delta\uparrow(1-\varepsilon)} \bar{H}_\delta(\boldsymbol{X}) - 4\gamma\right) + \gamma < \lim_{\delta\uparrow(1-\varepsilon)} \bar{H}_\delta(\boldsymbol{X}) - 2\gamma \quad (2.1.2)$$

and

$$\limsup_{n\to\infty} P_e(\mathcal{C}_n) \leq \varepsilon. \quad (2.1.3)$$

(2.1.2) implies that

$$\frac{1}{n}\log M_n \le \lim_{\delta\uparrow(1-\varepsilon)} \bar{H}_\delta(\boldsymbol{X}) - 2\gamma$$

for all sufficiently large n. Hence, for those n satisfying the above inequality and also by Lemma 2.4,

$$\begin{aligned} P_e(\mathscr{C}_n) &\ge \Pr\left[\frac{1}{n}h_{X^n}(X^n) > \frac{1}{n}\log M_n + \gamma\right] - e^{-n\gamma} \\ &\ge \Pr\left[\frac{1}{n}h_{X^n}(X^n) > \left(\lim_{\delta\uparrow(1-\varepsilon)} \bar{H}_\delta(\boldsymbol{X}) - 2\gamma\right) + \gamma\right] - e^{-n\gamma}. \end{aligned}$$

Therefore,

$$\begin{aligned} \limsup_{n\to\infty} P_e(\mathscr{C}_n) &\ge 1 - \liminf_{n\to\infty} \Pr\left[\frac{1}{n}h_{X^n}(X^n) \le \lim_{\delta\uparrow(1-\varepsilon)} \bar{H}_\delta(\boldsymbol{X}) - \gamma\right] \\ &> 1 - (1-\varepsilon) = \varepsilon, \end{aligned}$$

where the last inequality follows from (2.1.1). Thus, a contradiction to (2.1.3) is obtained. $\square$

Fixed-length codes for arbitrary sources

A few remarks are made based on the previous theorem.

- Note that as $\varepsilon = 0$,

$$T_0(\boldsymbol{X}) = \lim_{\delta \uparrow (1-\varepsilon)} \bar{H}_\delta(\boldsymbol{X}) = \bar{H}(\boldsymbol{X}).$$

  Hence, the minimum (asymptotic) lossless fixed-length source coding rate of any finite-alphabet source is $\bar{H}(\boldsymbol{X})$.

- Consider the special case where

  $-\dfrac{1}{n}\log P_{X^n}(X^n)$ converges in probability to a constant H (entropy rate),

  which holds for all *information stable* sources. In this case, both the inf- and sup-spectrums of $\boldsymbol{X}$ degenerate to a unit step function:

$$u(\theta) = \begin{cases} 1, & \text{if } \theta > H; \\ 0, & \text{if } \theta < H. \end{cases}$$

  Thus, $\bar{H}_\varepsilon(\boldsymbol{X}) = H$ for all $\varepsilon \in [0,1)$. Hence, general source coding theorem reduces to the conventional source coding theorem.

Fixed-length codes for arbitrary sources

– A source

$$\boldsymbol{X} = \left\{ X^n = \left(X_1^{(n)}, \ldots, X_n^{(n)} \right) \right\}_{n=1}^{\infty}$$

is said to be *information stable* if

$$H(X^n) = E\left[-\log P_{X^n}(x^n)\right] > 0 \text{ for all } n,$$

and

$$\lim_{n\to\infty} \Pr\left(\left| \frac{-\log P_{X^n}(x^n)}{H(X^n)} - 1 \right| > \varepsilon \right) = 0,$$

for every $\varepsilon > 0$.

– By the definition, any stationary-ergodic source with finite n-fold entropy is information stable; hence, it can be viewed as a generalized source model for stationary-ergodic sources.

Fixed-length codes for arbitrary sources

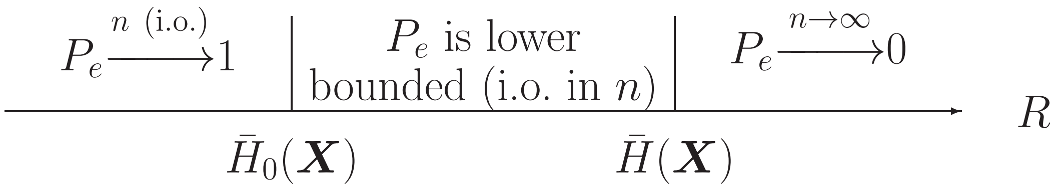
- If
$$-\frac{1}{n}\log P_{X^n}(X^n) \text{ converges in probability to a random variable } Z$$
whose cdf is $F_Z(\cdot)$, then the minimum achievable data compression rate subject to decoding error being no greater than ε is
$$T_\varepsilon(\boldsymbol{X}) = \lim_{\delta \uparrow (1-\varepsilon)} \bar{H}_\delta(\boldsymbol{X}) = \sup\left\{R \;:\; F_Z(R) < 1-\varepsilon\right\}.$$

Example 2.6 Consider a binary source $\boldsymbol{X}$ with each X^n is Bernoulli(Θ) distributed, where Θ is a random variable defined over $(0,1)$. By ergodic decomposition theorem (which states that any stationary source can be viewed as a mixture of stationary-ergodic sources) that
$$-\frac{1}{n}\log P_{X^n}(X^n) \text{ converges in probability to } h_b(\Theta),$$
where $h_b(x) := -x\log_2(x) - (1-x)\log_2(1-x)$. Consequently,
$$T_\varepsilon(\boldsymbol{X}) = \sup\{R \;:\; \Pr\{h_b(\Theta) \le R\} < 1-\varepsilon\}.$$

Fixed-length codes for arbitrary sources

- From the above example, or from Theorem 2.5, it shows that the *strong converse theorem* (which states that codes with rate below entropy rate will ultimately have decompression error approaching one) does not hold in general. However, one can always claim the *weak converse statement* for arbitrary sources.

Theorem 2.7 (weak converse theorem) For any block code sequence of ultimate rate $R < \bar{H}(\boldsymbol{X})$, the probability of block decoding failure P_e cannot be made arbitrarily small. In other words, there exists $\varepsilon > 0$ such that P_e is lower bounded by ε infinitely often in block length n.



Behavior of the probability of block decoding error as block length n goes to infinity for an *arbitrary* source $\boldsymbol{X}$.

Generalized AEP theorem

Theorem 2.8 (generalized asymptotic equipartition property for arbitrary sources) Fix $\varepsilon \in [0, 1)$. Given an arbitrary source $\boldsymbol{X}$, define

$$\mathcal{T}_n[R] := \left\{ x^n \in \mathcal{X}^n : -\frac{1}{n} \log P_{X^n}(x^n) \leq R \right\}.$$

Then for any $\delta > 0$, the following statements hold.

1.
$$\liminf_{n\to\infty} \Pr\left\{ \mathcal{T}_n[\bar{H}_\varepsilon(\boldsymbol{X}) - \delta] \right\} \leq \varepsilon \tag{2.1.4}$$

2.
$$\liminf_{n\to\infty} \Pr\left\{ \mathcal{T}_n[\bar{H}_\varepsilon(\boldsymbol{X}) + \delta] \right\} > \varepsilon \tag{2.1.5}$$

3. The number of elements in

$$\mathcal{F}_n(\delta; \varepsilon) := \mathcal{T}_n[\bar{H}_\varepsilon(\boldsymbol{X}) + \delta] \setminus \mathcal{T}_n[\bar{H}_\varepsilon(\boldsymbol{X}) - \delta],$$

denoted by $|\mathcal{F}_n(\delta; \varepsilon)|$, satisfies

$$|\mathcal{F}_n(\delta; \varepsilon)| \leq \exp\left\{ n(\bar{H}_\varepsilon(\boldsymbol{X}) + \delta) \right\}, \tag{2.1.6}$$

where the operation $\mathcal{A} \setminus \mathcal{B}$ between two sets $\mathcal{A}$ and $\mathcal{B}$ is defined by $\mathcal{A} \setminus \mathcal{B} := \mathcal{A} \cap \mathcal{B}^c$ with $\mathcal{B}^c$ denoting the complement set of $\mathcal{B}$.

Generalized AEP theorem

4. There exists $\rho = \rho(\delta) > 0$ and a subsequence $\{n_j\}_{j=1}^{\infty}$ such that

$$|\mathcal{F}_n(\delta;\varepsilon)| > \rho \cdot \exp\left\{n_j(\bar{H}_\varepsilon(\boldsymbol{X}) - \delta)\right\}. \quad (2.1.7)$$

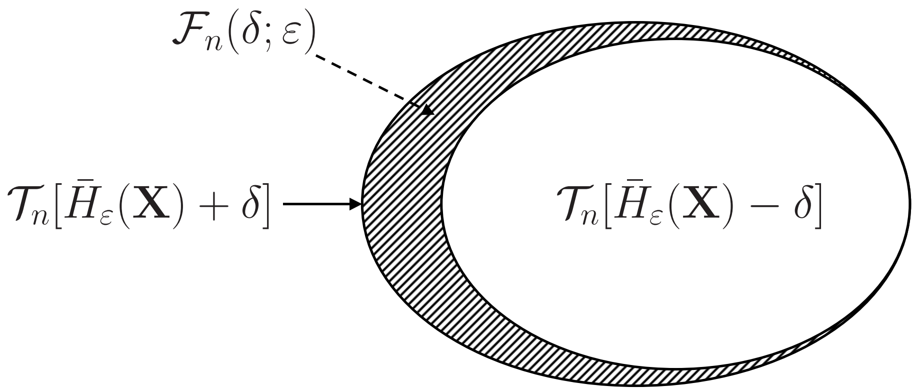


Illustration of generalized AEP Theorem. $\mathcal{F}_n(\delta;\varepsilon) := \mathcal{T}_n[\bar{H}_\varepsilon(\mathbf{X}) + \delta] \setminus \mathcal{T}_n[\bar{H}_\varepsilon(\mathbf{X}) - \delta]$ is the dashed region.

Generalized AEP theorem

- The set

$$\begin{aligned}\mathcal{F}_n(\delta;\varepsilon) &:= \mathcal{T}_n[\bar{H}_\varepsilon(\boldsymbol{X})+\delta] \setminus \mathcal{T}_n[\bar{H}_\varepsilon(\boldsymbol{X})-\delta] \\ &= \left\{x^n \in \mathcal{X}^n : \left|-\frac{1}{n}\log P_{X^n}(x^n) - \bar{H}_\varepsilon(\boldsymbol{X})\right| < \delta\right\} \\ &\qquad \bigcup \left\{x^n \in \mathcal{X}^n : -\frac{1}{n}\log P_{X^n}(x^n) = \bar{H}_\varepsilon(\boldsymbol{X})+\delta\right\}\end{aligned}$$

  is nothing but the weakly δ-typical set.

- $q_n := \Pr\{\mathcal{F}_n(\delta;\varepsilon)\} > 0$ infinitely often in n.

- $$|\mathcal{F}_n(\delta;\varepsilon)| \approx e^{n\bar{H}_\varepsilon(\boldsymbol{X})},$$

  and the probability of each sequence in $\mathcal{F}_n(\delta;\varepsilon)$ can be estimated by $q_n \cdot \exp\left\{-n\bar{H}_\varepsilon(\boldsymbol{X})\right\}$.

- In particular, if $\boldsymbol{X}$ is a stationary-ergodic source, then $\bar{H}_\varepsilon(\boldsymbol{X})$ is independent of $\varepsilon \in [0,1)$ and, $\bar{H}_\varepsilon(\boldsymbol{X}) = \underline{H}_\varepsilon(\boldsymbol{X}) = H$ for all $\varepsilon \in [0,1)$, where H is the source entropy rate

$$H = \lim_{n\to\infty}\frac{1}{n}E\left[-\log P_{X^n}(X^n)\right].$$

  In this case, the generalized AEP reduces to the conventional AEP.