Coding and Information Theory
Chapter 3
Entropy (B)

Xuejun Liang

This is the second lecture of chapter 3

Chapter 3: Entropy

3.1 Information and Entropy

3.2 Properties of the Entropy Function

3.3 Entropy and Average Word-length

3.4 Shannon-Fane Coding

3.5 Entropy of Extensions and Products

3.6 Shannon's First Theorem

3.7 An Example of Shannon's First Theorem

Quick Review of Last Lecture

$$H_r(S) = \sum_{i=1}^{q} p_i I_r(s_i) = \sum_{i=1}^{q} p_i \log_r \frac{1}{p_i} = -\sum_{i=1}^{q} p_i \log_r p_i$$

$$H(p) = -p \log p - \overline{p} \log \overline{p}.$$

$$H_r(S) = q \cdot \frac{1}{q} \log_r q = \log_r q\,.$$

Theorem 3.7: $H_r(S) \geq 0$, with equality if and only if $p_i = 1$ for some i (so that $p_j = 0$ for all $j \neq i$).

$$\sum_{i=1}^{q} x_i \log_r \frac{1}{x_i} \leq \sum_{i=1}^{q} x_i \log_r \frac{1}{y_i}\,,$$

Corollary 3.9

Theorem 3.10: If a source S has q symbols then $H_r(S) \leq log_r q$, with equality if and only if the symbols are equiprobable.

3.3 Entropy and Average Word-length

- Theorem 3.11
  - If C is any uniquely decodable $r\text{-ary}$ code for a source S, then $L(C) \geq H_r(S)$.
- The interpretation
  - Each symbol emitted by S carries $H_r(S)$ units of information, on average.
  - Each code-symbol conveys one unit of information, so on average each code-word of C must contain at least $H_r(S)$ code-symbols, that is, $L(C) \geq H_r(S)$.
  - In particular, sources emitting more information require longer code-words.

Proof of Theorem 3.11

$$\begin{aligned}
H_r(\mathcal{S}) &= \sum_{i=1}^{q} p_i \log_r\left(\frac{1}{p_i}\right) \\
&\leq \sum_{i=1}^{q} p_i \log_r\left(\frac{1}{y_i}\right) \\
&= \sum_{i=1}^{q} p_i \log_r(r^{l_i} K) \\
&= \sum_{i=1}^{q} p_i(l_i + \log_r K) \\
&= \sum_{i=1}^{q} p_i l_i + \log_r K \sum_{i=1}^{q} p_i \\
&= L(\mathcal{C}) + \log_r K \\
&\leq L(\mathcal{C})
\end{aligned}$$

$$\sum_{i=1}^{q} x_i \log_r \frac{1}{x_i} \leq \sum_{i=1}^{q} x_i \log_r \frac{1}{y_i},$$

Corollary 3.12

Given a source S with probabilities p_i, there is a uniquely decodable r-ary code C for S with $L(C) = H_r(S)$ if and only if $log_r(p_i)$ is an integer for each i , that is, each $p_i = r^{e_i}$ for some integer $e_i \leq 0$.

$$= \sum_{i=1}^{q} p_i \log_r\left(\frac{1}{p_i}\right)$$

$$\leq \sum_{i=1}^{q} p_i \log_r\left(\frac{1}{y_i}\right)$$

$$= L(C) + \log_r K$$

$$\leq L(C)$$

Corollary 3.12

Given a source S with probabilities p_i, there is a uniquely decodable r-ary code C for S with $L(C) = H_r(S)$ if and only if $log_r(p_i)$ is an integer for each i , that is, each $p_i = r^{e_i}$ for some integer $e_i \leq 0$.

Example 3.13

If S has q = 3 symbols s_i, with probabilities p_i = 1⁄4, 1⁄2, and 1⁄4 (see Examples 1.2 and 2.1).

$H_2(S) =$

A binary Huffman code C for S:

$L(C) =$

- Example 3.14
  - Let S have q = 5 symbols, with probabilities $p_i = 0.3, 0.2, 0.2, 0.2, 0.1$, as in Example 2.5.
    - In Example 3.3, $H_2(S)$ = 2.246, and
    - in Example 2.5, $L(C) = 2.3$, C binary Huffman code for S
  - By Theorem 2.8, every uniquely decodable binary code D for S satisfies $L(D) \geq 2.3 > H_2(S)$.
  - Thus no such uniquely decodable binary code D satisfies

$$L(D) = H_r(S)$$

  - What is the reason?

- Example 3.15
  - Let S have 3 symbols s_i, with probabilities $p_i = \frac{1}{2}, \frac{1}{2}, 0$.

  - Let S have 2 symbols s_i, with probabilities $p_i = \frac{1}{2}, \frac{1}{2}$.

Code Efficiency and Redundancy

- If C is an r-ary code for a source S, its efficiency is defined to be

$$\eta = \frac{H_r(\mathcal{S})}{L(\mathcal{C})}, \qquad (3.4)$$

  - So $0 \leq \eta \leq 1$ for every uniquely decodable code C for S
- The redundancy of C is defined to be $\bar{\eta} = 1 - \eta$.
  - Thus increasing redundancy reduces efficiency
- In Examples 3.13 and 3.14,
  - $\eta = 1$ and $\eta \approx 0.977$, respectively.

3.4 Shannon-Fano Coding

- Shannon-Fano codes
  - close to optimal, but easier to estimate their average word lengths.
- A Shannon-Fano code C for S has word lengths

$$l_i = \lceil \log_r(1/p_i) \rceil \,, \quad (3.5)$$

- So, we have

$$\log_r \frac{1}{p_i} \le l_i < 1 + \log_r \frac{1}{p_i} \,, \quad (3.6)$$

$$K = \sum_{i=1}^{q} r^{-l_i} \le \sum_{i=1}^{q} p_i = 1,$$

So Theorem 1.20 (Kraft's inequality) implies that there is an instantaneous r-ary code C for S with these word-lengths l_i

- Theorem 3.16
  - Every r-ary Shannon-Fano code C for a source S satisfies

$$H_r(\mathcal{S}) \leq L(\mathcal{C}) \leq 1 + H_r(\mathcal{S})$$

$$\log_r \frac{1}{p_i} \leq l_i < 1 + \log_r \frac{1}{p_i}, \quad (3.6)$$

- Corollary 3.17
  - Every optimal r-ary code D for a source S satisfies

$$H_r(\mathcal{S}) \leq L(\mathcal{D}) \leq 1 + H_r(\mathcal{S})$$

- Example 3.18
  - Let S have 5 symbols, with probabilities p_i= 0.3, 0.2, 0.2, 0.2, 0.1 as in Example 2.5
  - Compute Shannon-Fano code word length l_i, $L(C)$, η.
  - Compare with Huffman code.

Compute word length l_i of Shannon-Fano Code

$$l_i = \lceil \log_2(1/p_i) \rceil = \min\{n \in \mathbf{Z} \mid 2^n \geq 1/p_i\}$$