Coding and Information Theory
Chapter 3
Entropy (A)

Xuejun Liang

This is the first 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

The aim of this chapter

- Introduce the entropy function
  - which measures the amount of information emitted by a source
- Examine the basic properties of this function
- Show how it is related to the average word lengths of encodings of the source

3.1 Information and Entropy

- Define a number $I(s_i)$, for each $s_i \in S$, which represents
  - How much information is gained by knowing that S has emitted s_i
  - Our prior uncertainty as to whether s_i will be emitted and our surprise on learning that it has been emitted
- Therefore require that:
  1) $I(s_i)$ is a decreasing function of the probability p_i of s_i, with $I(s_i) = 0$ if $p_i = 1$;
  2) $I(s_i s_j) = I(s_i) + I(s_j)$, where S emits s_i and s_j consecutively and independently.

Entropy Function

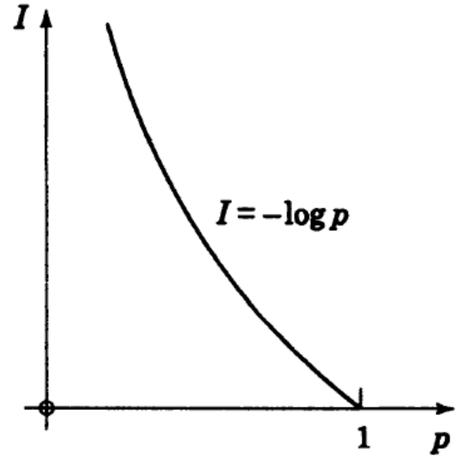


Figure 3.1

- We define

$$I(s_i) = -\log p_i = \log \frac{1}{p_i} \quad (3.1)$$

where $p_i = \Pr(s_i)$. So that I satisfies (1) and (2)

- Example 3.1
  - Let S be an unbiased coin, with s_1 and s_2 representing heads and tails. Then $I(s_1) =$? and $I(s_2) =$?

The r-ary Entropy of S

- The average amount of information conveyed by S (per source-symbol) is given by the function

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

- Called the r-ary entropy of S.
- Base r is often omitted

$$H(S) = \sum_{i=1}^{q} p_i \log \frac{1}{p_i} = -\sum_{i=1}^{q} p_i \log p_i$$

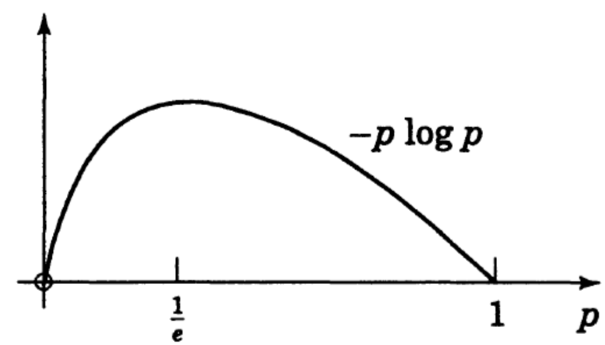


Figure 3.2

Example 3.2

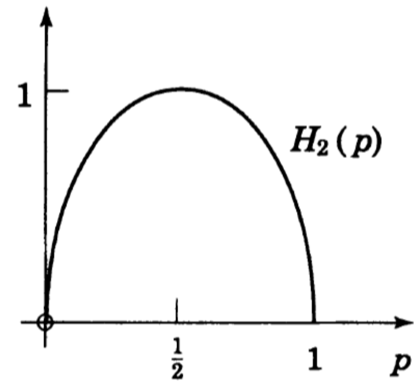


Figure 3.3

- Let S have q = 2 symbols, with probabilities p and 1 - p
- Let $\bar{p} = 1 - p$. Then

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

- $H(p)$ is maximal when p = ½
- Compute $H_2(p)$ when p = ½ and p = 2⁄3

Example 3.3

- If S has q = 5 symbols with probabilities
- p_i = 0.3, 0.2, 0.2, 0.2, 0.1, as in §2.2, Example 2.5,
- we find that $H_2(S) = 2.246$.

Examples (Cont.)

- If S has q equiprobable symbols, then $p_i = {}^1/_q$ for each i, so
$$H_r(S) = q \cdot \frac{1}{q} \log_r q = \log_r q \,.$$
- Example 3.4 and 3.5
  - Let $q = 5$, $H_2(S) = log_2 5 \approx 2.321$
  - Let $q = 6$, $H_2(S) = log_2 6 \approx 2.586$
- Example 3.6.
  - Using the known frequencies of the letters of the alphabet, the entropy of English text has been computed as approximately 4.03.

Compare average word-length of binary Huffman coding with entropy

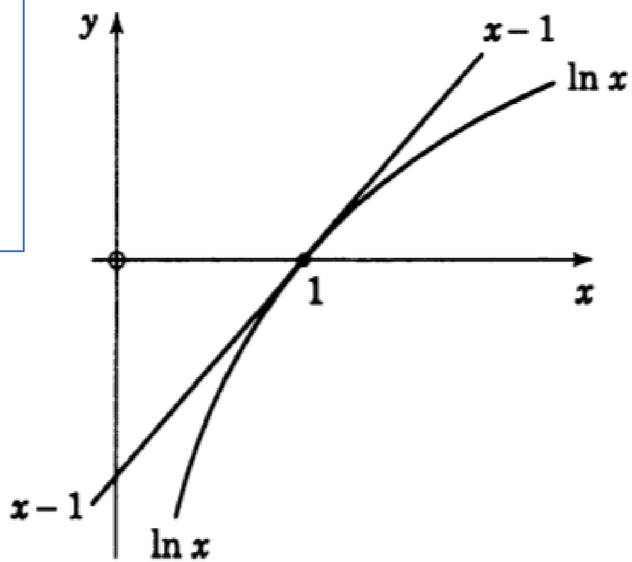
- As in Example 3.2 with $p = {}^{2}/_{3}$
  - $H_2(S) \approx 0.918$
  - $L(C^1) \approx 1, L(C^2)/2 \approx 0.944, L(C^3)/3 \approx 0.938$
- As in Example 3.3
  - $H_2(S) \approx 2.246$
  - $L(C^1) \approx 2.3$
- As in Example 3.4
  - $H_2(S) \approx 2.321$
  - $L(C^1) \approx 2.4$

3.2 Properties of the Entropy Function

- 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 $j \neq i$).

Lemma 3.8

For all $x > 0$ we have $\ln x \leq x - 1$, with equality if and only if $x = 1$.



Converting to some other base r, we have

$$log_r(x) \leq log_r(e) \cdot (x - 1)$$

with equality if and only if $x = 1$

Corollary 3.9

Let $x_i \geq 0$ and $y_i > 0$ for i = 1, ..., q, and let $\sum_i x_i = \sum_i y_i = 1$ (so (x_i) and (y_i) are probability distributions, with $y_i \neq 0$). Then

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

(that is, $\sum_i x_i \log(y_i/x_i) \leq 0$), with equality if and only if $x_i = y_i$ for all i.

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.

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