Source Coding

An alphabet of symbols $S_1, S_2, \cdots S_n$, each occurring with probability $p(S_1), p(S_2), \cdots p(S_n)$ is said to have an **entropy**

$$H = \sum_{k=1}^{n} p(S_k) \log_2 \frac{1}{p(S_k)}$$

What does entropy mean? How can we use it?

Q1 What does H mean?

Example Suppose N takes 32 values, $N=0, N=1, \cdots N=31$, all equally likely. Each outcome takes 5 bits to describe and has probability $\frac{1}{32}$.

$\log_2 \frac{1}{\frac{1}{32}} = \log_2 32 = 5$ bits = amount of information you gain from knowing value of N.

$\sum_{n=0}^{31} p(N=n)\log_2 \frac{1}{p(N=n)} = 5 =$ the **average** amount of information you gain from knowing the value of N, averaged over all the values N can take on.

Now consider the coarser events:

$$E_1 = \{k = 0,1,2,\cdots \text{ or } 15\} \quad p(E_1) = \tfrac{1}{2}$$
$$E_2 = \{k = 16,17,\cdots \text{ or } 23\} \quad p(E_2) = \tfrac{1}{4}$$
$$E_3 = \{k = 24,25,\cdots \text{ or } 27\} \quad p(E_3) = \tfrac{1}{8}$$
$$E_4 = \{k = 28,29,30 \text{ or } 31\} \quad p(E_4) = \tfrac{1}{8}$$

$\log_2 \frac{1}{p(E_1)} = \log_2 2 = 1 =$ amount of information you learned from knowing E_1 occurred.

$E_1 = \{\text{binary expression for } N \text{ is of form } 0XXXX\}$, **1 bit learned from E_1 occurrence**

$$\log_2 \frac{1}{p(E_2)} = \log_2 4 = 2$$

$E_2 = \Big\{\text{binary expression for } N \text{ is of form } \underbrace{10\text{XXX}}_{\text{2 bits known}}\Big\}$ $\log_2 \frac{1}{p(E_2)} = \log_2 4 = 2$ **bits learned**

$E_3 = \Big\{\text{binary expression for } N \text{ is of form } \underbrace{110\text{XX}}_{\text{3 bits known}}\Big\}$ $\log_2 \frac{1}{p(E_3)} = \log_2 8 = 3$ **bits learned**

$H = \sum_{k=1}^{4} p(E_k)\log_2 \frac{1}{p(E_k)} =$ the **average** # of bits of information gained from knowing what event occurred.

$$H = \frac{1}{2}\log_2 \frac{1}{\frac{1}{2}} + \frac{1}{4}\log_2 \frac{1}{\frac{1}{4}} + \frac{1}{8}\log_2 \frac{1}{\frac{1}{8}} + \frac{1}{8}\log_2 \frac{1}{\frac{1}{8}}$$

$$\left(\frac{1}{2}\right)(1) + \left(\frac{1}{4}\right)(2) + \left(\frac{1}{8}\right)(3) + \left(\frac{1}{8}\right)(3) = 1\tfrac{3}{4} \text{ bits}$$

Q2 What does H have to do with average code length L?

What would you guess?

Example

Use same probabilities as before, but now applied to symbols we wish to encode in bits:

S_1: $p(S_1) = \frac{1}{2}$

S_2: $p(S_2) = \frac{1}{4}$

S_3: $p(S_3) = \frac{1}{8}$

S_4: $p(S_4) = \frac{1}{8}$

Again, $H = 1\frac{3}{4}$ bits

You could use a 2-bit code for each

$S_1 \to 00$

$S_2 \to 01$

$S_3 \to 10$

$S_4 \to 11$

$$L = \sum_{k=1}^{4} p(S_k) l(S_n) = 2 \text{ bits}$$

Now try a variable-length code that gives shorter code words to more likely symbols. The Huffman coding procedure gives you, for example:

$S_1 \to 0$

$S_2 \to 10$

$S_3 \to 110$

$S_4 \to 111$

$$L = \sum_{k=1}^{4} p(S_k) l(S_n) = \left(\frac{1}{2}\right)(1) + \left(\frac{1}{4}\right)(2) + \left(\frac{1}{8}\right)(3) + \left(\frac{1}{8}\right)(3) = 1\frac{3}{4} \text{ bits}$$

Note that the average code length L is **now less than 2 !!!**

Note also that **L = H !!!**

Relation between H and L involves **Kraft Inequality** and **Gibbs Inequality.**

Definition: Prefix Code
A code is called a **prefix code** or an **instantaneous code** if no code word is a prefix for (i.e., the first several bits of) another code word.

Kraft Inequality

Let L_i be the length of the i-th codeword in an instantaneous code with n codewords. Then

$$\sum_{i=1}^{n} \frac{1}{2^{L_i}} \leq 1.$$

Tree Representation of Prefix Code

No codeword can be the parent of any other codeword

Gibbs Inequality

Let $\{p_k, 1 \le k \le m\}$ and $\{q_k, 1 \le k \le m\}$ be two probability distributions. Then

$$\sum_{k=1}^{n} p_k \ln \frac{1}{p_k} \le \sum_{k=1}^{n} p_k \ln \frac{1}{q_k}$$

with equality if and only if the two distributions are identical.

Note that this holds regardless of the base that is used for the logarithm. More interestingly, it also holds if each $q_k \ge 0$ but $\{q_k, 1 \le k \le n\}$ is not a probability distribution because

$$\sum_{k=1}^{n} q_k < 1.$$

Source Coding Theorem

For any alphabet of symbols $\{S_1, \cdots S_n\}$, any set of symbol probabilities $\{p_1, \cdots p_n\}$ and any binary prefix code for this alphabet

$$H \leq L.$$

Proof Use the Gibbs inequality, and let

$$q_k = \frac{1}{2^{L_k}},$$

and note from the Kraft inequality that

$$\sum_{k=1}^{n} q_k \leq 1.$$

Then

$$H = \sum_{k=1}^{n} p_k \log \frac{1}{p_k} \leq \sum_{k=1}^{n} p_k \log \frac{1}{q_k} = \sum_{k=1}^{n} p_k \log 2^{L_k} = \sum_{k=1}^{n} p_k L_k = L.$$

Note that this becomes an equality for a code with $p_k = \frac{1}{2^{L_k}}, k = 1, \cdots, n$, as in our previous example.