ENTROPY OF AN I.I.D. SOURCE

Let an I.I.D. source have \( M \) source letters that occur with probabilities \( p_1, p_2, \ldots, p_M \). Of course \( \sum_{\alpha=1}^{M} p_\alpha = 1 \).

The entropy of the source \( S \) is denoted \( H(S) \) and is defined as

\[
H(S) = \sum_{\alpha=1}^{M} p_\alpha \log_2 \frac{1}{p_\alpha} = -\sum_{\alpha=1}^{M} p_\alpha \log_2 p_\alpha
\]

\( H_a(S) = E \left[ \log_a \frac{1}{p_\alpha} \right] \)

The base of the logarithms is usually taken to be equal to 2. In that case \( H(S) \) is written as \( H_2(S) \) and is measured in units of "bits".

Other bases can be used. Since

\[
\log_a x = (\log_b x)(\log_b a) = (\log_b x) / (\log_b a)
\]

Then

\[
H_a(S) = H_2(S) \cdot \log_a b = H_2(S) / \log_a b
\]
A USEFUL THEOREM

Let \( p_1, p_2, \ldots, p_M \) be one set of probabilities and let \( p'_1, p'_2, \ldots, p'_M \) be another set of probabilities. (Note \( \sum_{i=1}^{M} p_i = 1 \) and \( \sum_{i=1}^{N} p'_i = 1 \))

**Theorem**

\[
\sum_{i=1}^{M} p_i \log \frac{1}{p_i} \leq \sum_{i=1}^{M} p_i \log \frac{1}{p'_i} \quad \text{with equality if } p_i = p'_i \quad \text{for } i = 1, 2, \ldots, M
\]

**Proof**

First note that \( \ln x \leq x - 1 \) with equality if \( x = 1 \)

\[
\sum_{i=1}^{M} p_i \log \frac{p_i}{p'_i} = \left( \sum_{i=1}^{M} p_i \ln \frac{p_i}{p'_i} \right) \log e
\]

\[
\leq (\log e) \sum_{i=1}^{M} p_i \left( \frac{p_i}{p'_i} - 1 \right)
\]

\[
= (\log e) \left( \sum_{i=1}^{M} p'_i - \sum_{i=1}^{M} p_i \right) = 0
\]

\[
\therefore \sum_{i=1}^{M} p_i \log \frac{1}{p_i} - \sum_{i=1}^{M} p_i \log \frac{1}{p'_i} = 0
\]

\[
\sum_{i=1}^{M} p_i \log \frac{1}{p_i} \leq \sum_{i=1}^{M} p_i \log \frac{1}{p'_i}
\]
MORE ON ENTROPY

1. For an equally likely i.i.d. source with $M$ source letters

$$H_2(S) = \log_2 M \quad (\text{or } H_0(S) = \log_2 M, \text{any } a_i)$$

2. For any i.i.d. source with $M$ source letters

$$0 \leq H_2(S) \leq \log_2 M \quad \text{follows from previous theorem with } p_i' = \frac{1}{M}, \forall i$$

$$\text{or } 0 \leq H_0(S) \leq \log_2 M, \text{any } a_i$$

3. Consider an i.i.d. source with $M$ source letters, $S$. If we consider encoding $m$ source letters at a time, this is an i.i.d. source with $M^m$ source letters. Call this the $m$th extension of the source and denote it by $S^m$. Then

$$H_2(S^m) = m \cdot H_2(S) \quad \text{(or } H_0(S^m) = m \cdot H_0(S), \text{any } a_i)$$

The proofs are omitted but are easy.
COMPUTATION OF ENTROPY (base 2)

**Example 2**

\[ M=2 \quad (p_1, p_2) = (0.9, 0.1) \]

\[ H_2(s) = 0.9 \log_2 \frac{1}{0.9} + 0.1 \log_2 \frac{1}{0.1} = 0.469 \text{ bits} \]

From before we gave Huffman codes for this source and extensions of this source for which

\[ \bar{L}_1 = 1 \]
\[ \bar{L}_{2/2} = 0.645 \]
\[ \bar{L}_{3/3} = 0.533 \]
\[ \bar{L}_{4/4} = 0.49 \]

Note that \( \bar{L}_m/m \geq H_2(s) \)

but as \( m \) gets larger \( \bar{L}_m/m \) is getting closer to \( H_2(s) \).

Holds in general:

\[ \Rightarrow \text{One can prove that for a binary Huffman code} \]

\[ H_2(s) \leq \bar{L}_m/m < H_2(s) + \frac{1}{m} \]
COMPUTATION OF ENTROPY (BASE 2)

EXAMPLE 2

\[ M = 3 \quad (p_1, p_2, p_3) = (0.5, 0.35, 0.15) \]

\[ H_2(\mathcal{S}) = 0.5 \log_2 0.5 + 0.35 \log_2 0.35 + 0.15 \log_2 0.15 = 1.44 \text{ bits} \]

But from before we gave codes for this source such that

1 symbol at a time \( \bar{L}_1 = 1.5 \)

2 symbols at a time \( \frac{\bar{L}_2}{2} = 1.46 \)
COMPUTATION OF ENTROPY (BASE 2)

Example 3

\[ M = 4 \quad (p_1, p_2, p_3, p_4) = \left( \frac{1}{2}, \frac{1}{4}, \frac{1}{8}, \frac{1}{8} \right) \]

\[ H_2(S) = \frac{1}{2} \log_2 \frac{1}{\frac{1}{2}} + \frac{1}{4} \log_2 \frac{1}{\frac{1}{4}} + \frac{1}{8} \log_2 \frac{1}{\frac{1}{8}} + \frac{1}{8} \log_2 \frac{1}{\frac{1}{8}} \]

\[ = 1.75 \text{ BITS} \]

But from before we gave the code

\[
\begin{align*}
\frac{1}{2} & \quad 0 \\
\frac{1}{4} & \quad 10 \\
\frac{1}{8} & \quad 110 \\
\frac{1}{8} & \quad 111
\end{align*}
\]

\[ \bar{L}_1 = 1.75 \]

In this special case, \( \bar{L}_1 = H_2(S) \). Thus one cannot improve on the efficiency of the code by encoding several letters at a time.
Significance of Entropy (Base 2)

For any U.D. code corresponding to the Nth extension of the IID source S, for each \( n = 1, 2, \ldots \)

\[
\frac{L_n}{N} \geq H(S)
\]

For a binary Huffman code corresponding to the Nth extension of the IID source S

\[
\frac{L_n}{N} \geq H_2(S) \quad \text{and} \quad \frac{L_n}{N} < H_2(S) + \frac{1}{N}
\]

But this implies

\[
\lim_{N \to \infty} \frac{L_n}{N} \to H(S)
\]
NON-BINARY CODE WORDS

The code symbols that make up the codewords can be from a higher order alphabet than 2.

**Example**

1-1-0 source \( \{A, B, C, D, E\} \)

<table>
<thead>
<tr>
<th>Source Symbols</th>
<th>Ternary Code</th>
<th>Quaternary Code</th>
<th>5-Level Code</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>B</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>C</td>
<td>20</td>
<td>20</td>
<td>2</td>
</tr>
<tr>
<td>D</td>
<td>21</td>
<td>30</td>
<td>3</td>
</tr>
<tr>
<td>E</td>
<td>22</td>
<td>31</td>
<td>4</td>
</tr>
</tbody>
</table>

Again, we are only interested in U.D. codes (where each concatenation of code words can be decoded in only one way).

A lower bound to the average code length of any U.D. code with \( r \)-levels is \( H_2(s) \) where

\[
H_2(s) = \sum_{\lambda=1}^{M} \rho_\lambda \log_2 \frac{1}{\rho_\lambda}
\]

For example, for a ternary code, the average length \( \bar{l}/m \), is no less than \( H_3(s) \).
KRAFT INEQUALITY

A necessary and sufficient condition for the construction of an instantaneous code with \( M \) code words of lengths \( l_1, l_2, \ldots, l_M \), where the code symbols take on \( r \) different values is that

\[
\sum_{i=1}^{M} \frac{r_i - l_i}{r_i} \leq 1
\]

Proof of Sufficiency. We construct an instantaneous code with these code word lengths. Let there be

\( M_j \) code words of length \( j \) for \( j = 1, 2, \ldots, l^*_k = \max l_i \).

Then

\[
\sum_{i=1}^{M} \frac{r_i - l_i}{r_i} = \sum_{j=1}^{l^*_k} m_j \frac{r - l}{r}
\]

Assume that

\[
\sum_{i=1}^{M} \frac{r_i - l_i}{r_i} = \sum_{j=1}^{l^*_k} m_j \frac{r - l}{r} \leq 1.
\]

Then

\[
\sum_{j=1}^{l^*_k} m_j r_i^{l_i-1} = r^{l^*_k}
\]

or

\[
m_k^* + m_{k-1}^* r + m_{k-2}^* r^2 + \ldots + m_1^* r^{l_1-1} \leq r^{l^*_k}
\]

But since \( m_k^* \geq 0 \) we then have

\[
0 \leq m_k^* \leq r^{l^*_k} - m_1^* r - m_2^* r^2 \ldots - m_{k-1}^* r
\]
But dividing by \( r \) and noting that \( m_{q-1} \geq 0 \) we have

\[
0 \leq m_q \leq r^{q-1} - m_{q-1} r^{q-2} - \ldots - m_{q-2} r^{q-3}
\]

Continuing we get

\[
0 \leq m_3 \leq r^3 - m_1 r^2 - m_2 r \\
0 \leq m_2 \leq r^2 - m_2 r \\
m_1 \geq r
\]

Note that if \( \sum_{i=1}^{m} r^{-i} \leq 1 \), then the \( m \) satisfy the above equations. Note that \( m_1 \leq r \). If \( m_1 < r \) we have \( (r-m_1) \) unused prefixes to form code words of length 1, for which the code words of length 1 are not prefixes. But \( m_2 \leq r^2 - m_2 r \). If \( m_2 < r^2 - m_2 r \) there are \( (r^2 - m_2 r - m_2) \) code words of length 3 which satisfy the prefix condition. Etc.

Proof of necessity: Follows from McMillan inequality,
McMillan Inequality

A NECESSARY AND SUFFICIENT CONDITION FOR THE EXISTENCE OF AN U.O.D. CODE WITH M CODE WORDS OF LENGTH \( l_1, l_2, \ldots, l_M \) WHERE THE CODE SYMBOLS TAKE ON "\( n \)" DIFFERENT VALUES IS

\[
\sum_{i=1}^{M} x^i \leq 1
\]

SKETCH OF PROOF OF NECESSITY

1. Assume \( \sum_{i=1}^{M} x^{-l_i} > 1 \) and a U.O.D. code exists

2. If \( \left( \sum_{i=1}^{M} x^{-l_i} > 1 \right) \), then \( \left[ \sum_{i=1}^{M} x^{-l_i} \right]^N \geq e^N \)

3. \( \left( \sum_{i=1}^{M} x^{-l_i} \right)^N = \sum_{k=m}^{m^\#} N_k x^{-k} \) \( \text{where} \ N_k = \# \text{of strings of } N \text{ code words that are all of length exactly } \ k \)

4. If the code is U.O.D. \( N_k \leq 2^k \). But then for a U.O.D. code

\[
\sum_{k=m}^{m^\#} N_k x^{-k} \leq \sum_{k=m}^{m^\#} 1 = M 2^k - N + 1
\]

which grows linearly with \( N \), not exponentially with \( N \).

Q.E.D.
Lower Bound for $\bar{L}$ for a U.D. code

$\bar{L} \geq H_n(S)$ where $\bar{L}$ = average length of U.D. code and $n = \#$ of symbols in code alphabet. $\bar{L} = H_n(S)$ if $\rho_i = n^{-d_i}$.

Proof: Let $\rho'_i = \frac{n^{-d_i}}{\sum_{j=1}^{M} n^{-d_j}}$. Note $\rho'_i > 0$ and $\sum_{i=1}^{M} \rho'_i = 1$.

From before

$H_n(S) = \sum_{i=1}^{M} \rho_i \log_2 \frac{1}{\rho_i} \leq \sum_{i=1}^{M} \rho_i \log_2 \rho_i$.

Then

$H_n(S) \leq \sum_{i=1}^{M} \rho_i d_i + \sum_{i=1}^{M} \rho_i \log_2 \left( \sum_{j=1}^{M} n^{-d_j} \right)$.

But for a U.D. code $\sum_{j=1}^{M} n^{-d_j} \leq 1$ so $\log_2 \left( \sum_{j=1}^{M} n^{-d_j} \right) = 0$.

$\therefore H_n(S) \leq \bar{L}$

Equality occurs if $\sum_{j=1}^{M} n^{-d_j} = 1$ and $\rho_i = \rho'_i$.

But both of these conditions hold if $\rho_i = n^{-d_i}, \forall i$ an integer.
HUFFMAN CODE AND ENTROPY

A U.D. coding scheme exists for a source with probabilities \( P_1, P_2, \ldots, P_m \) with average length \( L \) and \( m \) code symbols such that

\[
H_2(S) \leq L < H_2(S) + 1
\]

Proof

Choose \( \ell_a \) = the unique integer in the range

\[
\log_2 \frac{1}{P_a} \leq \ell_a < \log_2 \frac{1}{P_a} + 1 \quad \text{i.e.,} \quad \ell_a = \lceil \log_2 \frac{1}{P_a} \rceil
\]

Then

\[
\sum_{a=1}^{m} r_a - \ell_a \leq \sum_{a=1}^{m} P_a = 1 \quad \text{so a U.D. code exists with these lengths,}
\]

\[
\sum_{a=1}^{m} P_a \log_2 \frac{1}{P_a} \leq \sum_{a=1}^{m} P_a \ell_a < \sum_{a=1}^{m} P_a \log_2 \frac{1}{P_a} + \sum_{a=1}^{m} P_a
\]

\[H_2(S) \leq L < H_2(S) + 1 \quad \text{Q.E.D.}
\]

Then for an i.i.d. source, \( S \), a Huffman code exists with code alphabet \( S \# \# \) for the \( N \)th extension of this source such that

\[
H_2(S^N) \leq \frac{L_m}{m} < H_2(S^N) + 1
\]

Proof From Theorem above, a U.D. code exists for the \( N \)th extension of the source such that

\[
H_2(S^N) \leq L_m < H_2(S^N) + 1
\]

But \( H_2(S^N) = m H_2(S) \) and The Huffman code is at least as good as the U.D. code Q.E.D.
EXAMPLES of Codes with $H_a(2) < I < H_a(2) + 1$

**EX 1**

$r = 2$

<table>
<thead>
<tr>
<th>Prob</th>
<th>$\log_2 \frac{1}{p_a}$</th>
<th>$L_a = \lceil \log_2 \frac{1}{p_a} \rceil$</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>0.9</td>
<td>0.152 1</td>
</tr>
<tr>
<td>B</td>
<td>0.09</td>
<td>3.47 4</td>
</tr>
<tr>
<td>C</td>
<td>0.01</td>
<td>6.67 7</td>
</tr>
</tbody>
</table>

$H_a(2) = 0.516$

Note that $H_a(2) < I < H_a(2) + 1$

$0.516 < 1.33 < 1.516$

**Better Code (Actually Huffman Code)**

| A    | 0          |
| B    | 10         |
| C    | 11         |

$\bar{L} = 1.1$

**EX 2**

$r = 2$

<table>
<thead>
<tr>
<th>Prob</th>
<th>$\log_2 \frac{1}{p_a}$</th>
<th>$L_a = \lceil \log_2 \frac{1}{p_a} \rceil$</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>0.19</td>
<td>2.396 3</td>
</tr>
<tr>
<td>B</td>
<td>0.19</td>
<td>2.396 3</td>
</tr>
<tr>
<td>C</td>
<td>0.19</td>
<td>2.396 3</td>
</tr>
<tr>
<td>D</td>
<td>0.19</td>
<td>2.396 3</td>
</tr>
<tr>
<td>E</td>
<td>0.19</td>
<td>4.322 5</td>
</tr>
<tr>
<td>F</td>
<td>0.05</td>
<td></td>
</tr>
</tbody>
</table>

$H_a(2) = 2.492$

Note

**Huffman Code**

| A    | 00         |
| B    | 01         |
| C    | 100        |
| D    | 101        |
| E    | 110        |
| F    | 111        |

$\bar{L} = 2.62$
**Ex 3** \( n = 3 \)

<table>
<thead>
<tr>
<th>Prob</th>
<th>( \log_3 \frac{1}{p_i} )</th>
<th>( L_i = \sum \log_3 \frac{1}{p_i} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>0.19</td>
<td>1.51</td>
</tr>
<tr>
<td>B</td>
<td>0.19</td>
<td>1.51</td>
</tr>
<tr>
<td>C</td>
<td>0.19</td>
<td>1.51</td>
</tr>
<tr>
<td>D</td>
<td>0.19</td>
<td>1.51</td>
</tr>
<tr>
<td>E</td>
<td>0.19</td>
<td>1.51</td>
</tr>
<tr>
<td>F</td>
<td>0.05</td>
<td>2.78</td>
</tr>
</tbody>
</table>

\[ H_3(3) = 1.57 \]

**Huffman Code**

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Code</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>0</td>
</tr>
<tr>
<td>B</td>
<td>10</td>
</tr>
<tr>
<td>C</td>
<td>11</td>
</tr>
<tr>
<td>D</td>
<td>12</td>
</tr>
<tr>
<td>E</td>
<td>20</td>
</tr>
<tr>
<td>F</td>
<td>21</td>
</tr>
</tbody>
</table>

\[ L = 1.81 \]
Properties of an Optimal (compact) Binary Code

We only need to consider instantaneous codes!!

1. If $P_i < P_j$, then $l_i > l_j$

Proof: Otherwise switching code words will reduce $I$

2. There is no single code word of length $l_i$'s max $l_i$

Proof: If there were shorter it by one digit, it will still not be a prefix of any other code word and will shorten $I$.

3. Of the code words of length $l_i$, they occur in pairs in which the code words in each pair agree in all but the last digit.

Proof: If not, shorten the code word $P_i$ which is not the case by one digit and it will not be the prefix of any other code word. This will shorten $I$. 
Proof of Optimality of Binary Huffman Codes.

Theorem:

Consider a code $C_j$ of length $j$, where $j > 1$. Suppose there exists a code $C_{j-1}$ that is shorter than $C_j$. Let $L_{j-1}$ be the average length of $C_{j-1}$ and $L_j$ be the average length of $C_j$.

**Case 1:**
If $L_{j-1} < L_j$, then $C_j$ cannot be optimal, as there exists a shorter code $C_{j-1}$.

**Case 2:**
If $L_{j-1} > L_j$, then $C_j$ cannot be optimal, as there exists a longer code $C_{j-1}$ that is still shorter than $C_j$.

Therefore, for $C_j$ to be optimal, $L_{j-1} = L_j$.

Proof:

Suppose there were a better code at $(j-1)$. Call it $C_{j-1}'$. Its average length $L_{j-1}' < L_{j-1}$. But the two code words with probabilities $p_{a,0}$ and $p_{a,1}$ are identical except for the last digit. Form a new code $C_j'$ at $j$ that has the identical prefix as the code word for $p_{a,0}$. This code will have average length $L_j' = L_{j-1}' + (p_{a,0} + p_{a,1})$ so that $L_{j-1}' < L_{j-1}$. But this can't be the case if $C_j$ was optimal. Q.E.D.