Hartley’s Information Measure

- Messages are strings of characters from a fixed alphabet.
- The amount of information contained in a message should be a function of the total number of possible messages.
- If you have an alphabet with \( s \) symbols, then there are \( s^\ell \) messages of length, \( \ell \).
- The amount of information contained in two messages should be the sum of the information contained in the individual messages.
Hartley’s Information Measure (contd.)

• The amount of information in $\ell$ messages of length one should equal the amount of information in one message of length $\ell$.

It is clear that the only function which satisfies these requirements is the log function:

$$\ell \log(s) = \log(s^{\ell}).$$

If the base of the logarithm is two, then the unit of information is the *bit*. 
Shannon’s Information Measure

Let $X$ be a discrete r.v. with $n$ outcomes, $\{x_1, ..., x_n\}$. The probability that the outcome will be $x_i$ is $p_X(x_i)$. The information contained in a message about the outcome of $X$ is:

$$- \log p_X(x_i).$$

The avg. information or entropy of a message about the outcome of $X$ is:

$$H_X = - \sum_{i=1}^{n} p_X(x_i) \log p_X(x_i).$$
Example

Let $X$ be a discrete r.v. with two outcomes, $\{x_1, x_2\}$. The probability that the outcome will be $x_1$ is $\theta$ and the probability that the outcome will be $x_2$ is $1 - \theta$. The avg. information contained in a message about the outcome of $X$ is:

$$H_X = -\theta \log(\theta) - (1 - \theta) \log(1 - \theta).$$

We observe that the avg. information is maximized when $\theta = 1 - \theta = \frac{1}{2}$, in which case $H_X = 1$ bit.
Joint Information

Let $X$ be a discrete r.v. with outcomes, $\{x_1, \ldots, x_n\}$ and let $Y$ be a discrete r.v. with outcomes, $\{y_1, \ldots, y_m\}$. The probability that the outcome of $X$ is $x_i$ and the outcome of $Y$ is $y_j$ is $p_{XY}(x_i, y_j)$. The amount of information contained in a message about the outcome of $X$ and $Y$ is:

$$- \log p_{XY}(x_i, y_j).$$

The avg. information or entropy of a message about the outcome of $X$ and $Y$ is:

$$H_{XY} = - \sum_{i=1}^{n} \sum_{j=1}^{m} p_{XY}(x_i, y_j) \log p_{XY}(x_i, y_j).$$
Figure 1: $H_X = -\theta \log(\theta) - (1 - \theta) \log(1 - \theta)$. 
Properties of Shannon’s Measure

• $H_X$ is continuous in the $p_X(x_i)$.

• $H_X$ is symmetric. That is, $H_X = H_Y$ when $p_Y(x_1) = p_X(x_2)$ and $p_Y(x_2) = p_X(x_1)$. More generally, $H_X$ is invariant under permutation of the distribution function, $p_X$.

• $H_X$ is additive. That is, when $X$ and $Y$ are independent r.v.’s, then $H_{XY} = H_X + H_Y$.

• $H_X$ is maximum when all of the $p_X(x_i)$’s are equal.

• $H_X$ is minimum when one of the $p_X(x_i)$’s equals one.
Additivity Example

Let $X$ and $Y$ be fair dice. The avg. amount of information contained in a message about the outcome of $X$ and $Y$ is:

$$H_{XY} = - \sum_{i=1}^{6} \sum_{j=1}^{6} \frac{1}{36} \log \frac{1}{36} \approx 5.16 \text{ bits.}$$

The avg. amount of information contained in a message about the outcome of $X$ is:

$$H_{X} = - \sum_{i=1}^{6} \frac{1}{6} \log \frac{1}{6} \approx 2.58 \text{ bits.}$$

Since $H_{X} = H_{Y}$, it follows that $H_{X} + H_{Y} \approx 5.16 \text{ bits.}$
Symmetry Example

- Let $X$ be a discrete r.v. with outcomes, \{A, G, C, T\}, which occur with probabilities, \{\frac{1}{4}, \frac{1}{8}, \frac{1}{8}, \frac{1}{2}\}.

- Let $Y$ be a discrete r.v. with outcomes, \{♣, ♠, ♦, ♥\}, which occur with probabilities, \{\frac{1}{4}, \frac{1}{8}, \frac{1}{2}, \frac{1}{8}\}.

- The avg. amount of information contained in a message about the outcome of $X$ is:

  \[
  H_X = -\frac{1}{4}\log\frac{1}{4} - \frac{1}{8}\log\frac{1}{8} - \frac{1}{8}\log\frac{1}{8} - \frac{1}{2}\log\frac{1}{2}
  = 1.75 \text{ bits}
  \]
Symmetry Example (contd.)

- The avg. amount of information contained in a message about the outcome of $Y$ is:

$$H_Y = -\frac{1}{4}\log\frac{1}{4} - \frac{1}{8}\log\frac{1}{8} - \frac{1}{2}\log\frac{1}{2} - \frac{1}{8}\log\frac{1}{8}$$

$$= 1.75 \text{ bits}$$
Theorem 1.1

Let $X$ be a discrete r.v. with $n$ outcomes, $\{x_1, \ldots, x_n\}$. The probability that the outcome will be $x_i$ is $p_X(x_i)$. Then

- $H_X \leq \log n$ with $H_X = \log n$ if and only if for all $i$ it is true that $p_X(x_i) = 1/n$.

- $H_X \geq 0$ with $H_X = 0$ if and only if there exists a $k$ such that $p_X(x_k) = 1$. 
Theorem 1.1 (contd.)

Proof:

\[ H_X - \log n = \]

\[ = - \sum_{i=1}^{n} p_X(x_i) \log p_X(x_i) - \log n = \]

\[ = - \sum_{i=1}^{n} p_X(x_i) \log p_X(x_i) - \sum_{i=1}^{n} p_X(x_i) \log n = \]

\[ = - \sum_{i=1}^{n} p_X(x_i) (\log p_X(x_i) + \log n) = \]

\[ = \sum_{i=1}^{n} p_X(x_i) \log \left( \frac{1}{np_X(x_i)} \right). \]
Theorem 1.1 (contd.)

From the inequality $\ln a \leq a - 1$ and the fact that $\log a = \ln a / \ln 2$:

\[
\begin{align*}
\ln a & \leq (a - 1) \\
\ln a / \ln 2 & \leq (a - 1) / \ln 2 \\
\log a & \leq (a - 1) / \ln 2 \\
\log a & \leq (a - 1) \ln e / \ln 2 \\
\log a & \leq (a - 1) \log e
\end{align*}
\]
Figure 2: $\ln a \leq a - 1$. 

\[
\text{line 1} \\
\text{line 2}
\]
Theorem 1.1 (contd.)

Using this result in the expression for $H_X - \log n$ yields:

$$H_X - \log n = \sum_{i=1}^{n} p_X(x_i) \log \left( \frac{1}{np_X(x_i)} \right)$$

$$\leq \sum_{i=1}^{n} p_X(x_i) \left( \frac{1}{np_X(x_i)} - 1 \right) \log e$$

$$\leq \left( \sum_{i=1}^{n} \frac{1}{n} - \sum_{i=1}^{n} p_X(x_i) \right) \log e$$

$$\leq \left( \frac{1}{n} - 1 \right) \log e$$

$$\leq 0$$
Theorem 1.1 (contd.)

This proves that $H_X \leq \log n$. To prove that $H_X \geq 0$, we observe that:

• $\forall i \ p_X(x_i) \geq 0$
• $\forall i \ -\log p_X(x_i) \geq 0$

It follows that:

$$-\sum_{i=1}^{n} p_X(x_i) \log p_X(x_i) \geq 0.$$
Maximum Entropy

Let $X$ be a r.v. with outcomes, $\{x_1, \ldots, x_n\}$. These outcomes occur with probability, $p_X(x_i) = 1/n$ for all $i$. The avg. information contained in a message about the outcome of $X$ is:

$$H_X = - \sum_{i=1}^{n} p_X(x_i) \log p_X(x_i)$$

$$= - \sum_{i=1}^{n} \frac{1}{n} \log \frac{1}{n}$$

$$= - \left( \frac{1}{n} \log \frac{1}{n} \right) \sum_{i=1}^{n} 1$$

$$= - \left( \frac{1}{n} \log \frac{1}{n} \right) n = - \log \frac{1}{n}$$

$$= \log n$$
Maximum Entropy Example

Let $X$ be a discrete r.v. with outcomes, \{A, G, C, T\}. These outcomes occur with probabilities, $\{\frac{1}{4}, \frac{1}{4}, \frac{1}{4}, \frac{1}{4}\}$. The avg. amount of information contained in a message about the outcome of $X$ is:

$$H_X = -\frac{1}{4} \log \frac{1}{4} - \frac{1}{4} \log \frac{1}{4} - \frac{1}{4} \log \frac{1}{4} - \frac{1}{4} \log \frac{1}{4}$$

$$= 2 \text{ bits}$$

The genome of the bacterium, *E. coli*, is a DNA molecule consisting of $4 \times 10^6$ base pairs. The maximum amount of information stored in the *E. coli* genome is therefore $8 \times 10^6$ bits.
Minimum Entropy Example

Let $X$ be a discrete r.v. with outcomes, \{A, G, C, T\}. These outcomes occur with probabilities, \{0, 1, 0, 0\}. The avg. amount of information contained in a message about the outcome of $X$ is:

$$H_x = -0 \log 0 - 1 \log 1 - 0 \log 0 - 0 \log 0$$
$$= 0 \text{ bits}.$$