1. \( p(x=0) = 1-p \quad 0 \leq p \leq 1 \)
\( p(x=1) = p \)

(a) \( h(p) = H_2(x) = -p(x=0) \log_2 p(x=0) - p(x=1) \log_2 p(x=1) \)
\[ = -(1-p) \log_2 (1-p) - p \log_2 p \]

(b) Observe that \( h(1-p) = h(p) \)
\[ \therefore h(p) \text{ is symmetric about } p = \frac{1}{2}. \]

(c) \( h(p) = -(1-p) \frac{\log(1-p)}{\log 2} - p \frac{\log p}{\log 2} \)

\[ h'(p) = \frac{4}{dp} h(p) \]
\[ = -\frac{(1-p)}{\log 2} \frac{1}{1-p} \cdot (-1) + \frac{\log(1-p)}{\log 2} - \frac{p}{\log 2} \cdot \frac{1}{p} - \frac{\log p}{\log 2} \]
\[ = \log_2 \left( \frac{1-p}{p} \right) \]

2. \( X = (X, x) \)
\( Y = (Y, y) \)

Define \( Z = (Z, z) \)
where \( Z = X \times Y \)
\( Y(Z) = p(x) q(y) \quad \text{for } Z = (x, y) \)

\[ H_2(Z) = - \sum_{z \in Z} Y(z) \log_2 Y(z) \]
\[ = - \sum_{x \in X} \sum_{y \in Y} p(x) q(y) \log_2 (p(x) q(y)) \]
\[
= \sum_{x \in \mathbb{X}} \sum_{y \in \mathbb{Y}} p(x) q(y) \log_2 p(x) - \sum_{x \in \mathbb{X}} \sum_{y \in \mathbb{Y}} p(x) q(y) \log_2 q(y)
\]

\[
= \sum_{x \in \mathbb{X}} p(x) \log_2 p(x) \sum_{y \in \mathbb{Y}} q(y) \quad \text{=} \quad \sum_{y \in \mathbb{Y}} q(y) \log_2 q(y) \sum_{x \in \mathbb{X}} p(x)
\]

\[
= H_2(X) + H_2(Y)
\]

\[\therefore H_2(Z) = H_2(X) + H_2(Y)\]

\[Z = (Z, I), \quad Z = \mathbb{X} \times \mathbb{Y}\]

Define \(X = (X, \mathbb{P})\), \(Y = (Y, \mathbb{Q})\) with

\[p(x) \triangleq \sum_{y \in \mathbb{Y}} r(x, y), \quad x \in \mathbb{X}\]

\[q(y) \triangleq \sum_{x \in \mathbb{X}} r(x, y), \quad y \in \mathbb{Y}\]

(a) No. \(Z = (Z, I)\) can be determined from \(X = (X, \mathbb{P})\) and \(Y = (Y, \mathbb{Q})\) only when

\[r(x, y) = p(x) \cdot q(y), \quad x \in \mathbb{X}, \quad y \in \mathbb{Y}\]

(b) \(H_2(Z) = \sum_{x \in \mathbb{X}} \sum_{y \in \mathbb{Y}} r(x, y) \log_2 r(x, y)\)

\[= \sum_{x \in \mathbb{X}} \sum_{y \in \mathbb{Y}} r(x, y) \log_2 (p(x) \cdot q(y|x)) \quad \text{(with some slight abuse of notation to represent the conditional distribution)}\]

\[= \sum_{x \in \mathbb{X}} \sum_{y \in \mathbb{Y}} r(x, y) \log_2 p(x) - \sum_{x \in \mathbb{X}} \sum_{y \in \mathbb{Y}} r(x, y) \log_2 q(y|x)\]

\[= \sum_{x \in \mathbb{X}} \left( \sum_{y \in \mathbb{Y}} r(x, y) \right) \log_2 p(x) - \sum_{x \in \mathbb{X}} p(x) \sum_{y \in \mathbb{Y}} q(y|x) \log_2 q(y|x)\]
\[ H_2(Z) = H_2(X) - \sum_x p(x) \sum_y q(y|x) \log_2 q(y|x) \quad (1) \]

For arbitrary pmfs \( p(\cdot), q(\cdot) \), we know

\[ D(p \| q) \geq 0, \text{ with } D(p \| q) = 0 \iff p = q \]

Hence, \( \forall x \)

\[ D(q(\cdot|x) \| q(\cdot)) \geq 0 \]

i.e.

\[ \sum_y q(y|x) \log_2 \frac{q(y|x)}{q(y)} \geq 0 \]

i.e.

\[ -\sum_y q(y|x) \log_2 q(y|x) \leq -\sum_y q(y|x) \log_2 q(y) \quad (2) \]

Using (2) in (1),

\[ H_2(Z) \leq H_2(X) - \sum_{x,y} q(x,y) \log_2 \left( \sum_{y \in q(y)} \frac{q(x,y)}{q(y)} \log_2 q(y) \right) \]

\[ = H_2(X) - \sum_{y \in q(y)} \left( \sum_{x \in \mathcal{X}} q(x,y) \log_2 q(y) \right) \]

\[ = H_2(X) + H_2(Y) \]

Intuitively, \( H_2(X) \) is the entropy of first component of \( Z \), while \( H_2(Y) \) is the entropy of the second component alone. The entropy of \( Z \) can not exceed the sum of the entropies of each component.
(c) Note that
\[ H_2(2) = H_2(x) + H_2(y) \] with equality only when
\[ \forall x, \quad - \sum_y q(y|x) \log_2 q(y|x) = - \sum_y q(y|x) \log_2 q(y) \quad \text{in } \mathbb{R}^2. \]

\[ \iff \forall x, \quad D(q(y|x) \mid \mid q(y)) = 0 \]

\[ \iff \forall x, \quad q(y|x) = q(x) \]

or \[ q(y|x) = q(y), \quad y \in Y. \]

In other words, \( X \) and \( Y \) are independent and
\[ q(x, y) = p(x) \cdot q(y), \quad x \in X, \ y \in Y. \]

In Problem 2, this condition is satisfied and hence we obtain
\[ H_2(2) = H_2(x) + H_2(y). \]

4. Let \( X \) denote the outcome in a single trial.
\[ P(X = "H") = P(X = "T") = \frac{1}{2}. \]

\[ \therefore H_2(x) = \log_2 2 = 1 \text{ bit} \]

To calculate entropy for 10 trials, look at the extended source of order 10.
\[ H(x^{10}) = 10 \cdot H(x) = 10 \text{ bits.} \]

(\text{Since } H(x^n) = n \cdot H(x)).
Observe that the value of the sum $\sum_{x \in \mathcal{X}} 2^{-l(x)}$ decreases as the value of $l(x)$ increases for any $x \in \mathcal{X}$.

Hence, for any finite set $\mathcal{X}$, it is always possible to find $\{l(x), x \in \mathcal{X}\}$ such that

$$\sum_{x \in \mathcal{X}} 2^{-l(x)} \leq 1$$

by choosing the lengths to be large enough.

Thus, Kraft inequality says there exists a prefix code $C : \mathcal{X} \to \mathbb{B}^*$ such that

$$l_C(x) = l(x), \ x \in \mathcal{X}$$

i.e.,

$$\sum_{x \in \mathcal{X}} 2^{-l_C(x)} \leq 1 \quad \text{--- (1)}$$

Now, if we want to construct a prefix code $C_\lambda : \mathcal{X} \to \mathbb{B}^*$ such that

$$\sum_{x \in \mathcal{X}} 2^{-l_{C_\lambda}(x)} \leq \lambda, \ \lambda \in (0, 1) \quad \text{--- (2)}$$

we just need to pad the code $C : \mathcal{X} \to \mathbb{B}^*$ (from (1)) with the required no. of $0$'s (or $1$'s). Note that such padding will not affect the prefix condition.

For e.g., if we pad all codewords in $C : \mathcal{X} \to \mathbb{B}^*$ with "a" zeroes to get $C_\lambda : \mathcal{X} \to \mathbb{B}^*$, then

$$\sum_{x \in \mathcal{X}} 2^{-l_{C_\lambda}(x)} = \sum_{x \in \mathcal{X}} 2^{-(l(x)+a)} = 2^{-a} \cdot \sum_{x \in \mathcal{X}} 2^{-l(x)}$$

$$\leq 2^{-a} \quad \text{(from (1))}$$
by choosing \( a \) s.t.

\[ 2^a \leq \lambda \quad \text{i.e.} \quad a \geq -\log_2 \lambda \]

we can ensure condition (2) is satisfied.

Hence, this is \( \text{NOT} \) an interesting fact.

(b) For a given alphabet \( \mathcal{X} \), if we can find \( \{ l(x), x \in \mathcal{X} \} \) s.t.

\[ \sum_{x \in \mathcal{X}} 2^{-l(x)} = 1 \quad (3) \]

then by Kraft inequality there exists at least one prefix code which satisfies the stated condition. So, the question is really whether we can find \( \{ l(x), x \in \mathcal{X} \} \) satisfying (3) for any given \( \mathcal{X} \). Obviously, this depends only on the no. of symbols in \( \mathcal{X} \), i.e. \( |\mathcal{X}| \).

For \( |\mathcal{X}| = 1 \), i.e. there only a single symbol, there is no need for any code and (3) is trivially satisfied by zero codeword length.

For \( |\mathcal{X}| = 2 \), the codeword length \( \{ 1, 1 \} \) satisfy (3).

We will now use an induction argument to show such lengths can be found for any (finite) alphabet \( \mathcal{X} \).

Suppose there exist \( \{ l(x), x \in \mathcal{X} \} = \{ l_1, \ldots, l_k \} \) satisfying (3) for \( |\mathcal{X}| = k \),

we want to show there exists \( \{ l_1, \ldots, l_{k+1} \} \) that satisfy (3) for \( |\mathcal{X}| = k+1 \). Simply choose \( l'_1 = l_1, \ldots, l'_{k-1} = l_{k-1}, l'_k = l_k + 1 \).

Then,

\[ \sum_{i=1}^{k+1} 2^{-l'_i} = \sum_{i=1}^{k-1} 2^{-l_i} + 2^{-l_k} \cdot 2^{-1} = \sum_{i=1}^{k} 2^{-l_i} = 1 \quad \text{(by assumption)}. \]

\( \therefore \) it is indeed \( \text{TRUE} \) that there exists at least one prefix code \( c : \mathcal{X} \rightarrow \{0,1\}^* \) s.t.

\[ \sum_{x \in \mathcal{X}} 2^{-l(c(x))} = 1. \]
Given lengths: 1, 3, 3, 4, 4.

Kraft inequality satisfied?

\[ 2^{-1} + (2^{-2} \times 3) + (2^{-4} \times 2) = 1 \quad \text{YES!} \]

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>00</td>
</tr>
<tr>
<td></td>
<td>101</td>
</tr>
<tr>
<td></td>
<td>110</td>
</tr>
<tr>
<td></td>
<td>1100</td>
</tr>
<tr>
<td></td>
<td>1101</td>
</tr>
</tbody>
</table>

"there exists a prefix code with given codeword lengths.

Given lengths: 2, 2, 3, 3, 4, 4, 5, 5

Kraft inequality satisfied?

\[ (2^2 \times 2) + (2^3 \times 2) + (2^4 \times 2) + (2^5 \times 2) = \frac{15}{16} \leq 1 \quad \text{YES!} \]

"there exists a prefix code with given codeword lengths.

The prefix code contains

<p>| |</p>
<table>
<thead>
<tr>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
</tr>
<tr>
<td>10</td>
</tr>
<tr>
<td>110</td>
</tr>
</tbody>
</table>

So, the only prefix for the other

codewords is 111.
Since we need the maximal number of codewords of length 5, we may include all those with prefix 111:

\[
\begin{array}{c}
11100 \\
11101 \\
11110 \\
11111
\end{array}
\]
as they satisfy the prefix condition.

\[\text{maximal no. of codewords of length 5} = 4\]

and total codewords = \[3 + 4 = 7\].

Since each symbol in \[\mathcal{E}\] is assigned a unique codeword \[|\mathcal{E}| = 7\].

9. To begin with, we have 16 possibilities for \(X\):

\[0, 1, 2, \ldots, 15\]

By asking each YES/NO question, we will shorten the no. of possibilities for \(X\) by \(\frac{1}{2}\).

For eg: we ask whether \(X \leq 7\)? If YES, we ask whether \(X \leq 3\)? and so on...

\[
\begin{array}{c}
X \leq 7? \\
\downarrow \text{Yes} \quad \downarrow \text{No} \\
X \leq 11? \\
\downarrow \text{Yes} \quad \downarrow \text{No} \\
X \leq 5? \\
\downarrow \text{Yes} \quad \downarrow \text{No} \\
X \leq 3? \\
\downarrow \text{Yes} \quad \downarrow \text{No} \\
X = 0? \\
\downarrow \text{Yes} \quad \downarrow \text{No} \\
X = 1? \quad X = 2? \quad X = 3? \quad X = 4? \quad X = 5? \quad X = 6? \quad X = 7?
\end{array}
\]

Similarly on this side.
X can be determined with 4 questions.

Connection to entropy: Since all values are equally probable

\[ P(X = i) = \frac{1}{16}, \quad i = 0, 1, \ldots, 15 \]

\[ \therefore H_2(X) = \log_2 16 = 4 \]

\[ \therefore H_2(Y) = - \sum_{y \in Y} q(y) \log_2 q(y) \]

\[ = - \sum_{y \in Y} \left( \sum_{x \in X; q(x) = y} p(x) \right) \log_2 \left( \sum_{x \in X; q(x) = y} p(x) \right) \quad \text{(1)} \]

Observe that, in general, for \( a_i > 0, \quad i = 1, \ldots, n \)

\[ \left( \sum_{i=1}^{n} a_i \right) \log_2 \left( \sum_{i=1}^{n} a_i \right) = \left( \sum_{i=1}^{n} a_i \log_2 \left( \sum_{i=1}^{n} a_i \right) \right) \]

\[ \geq \sum_{i=1}^{n} a_i \log_2 a_i \]

\[ \therefore - \left( \sum_{i=1}^{n} a_i \right) \log_2 \left( \sum_{i=1}^{n} a_i \right) \leq - \sum_{i=1}^{n} a_i \log_2 a_i \]

with equality only when the summation involves only one term, i.e. \( n = 1 \).

Using this in (1),

\[ H_2(Y) \leq - \sum_{y \in Y} \sum_{x \in X; q(x) = y} p(x) \log_2 p(x) \quad \text{(2)} \]

\[ = - \sum_{x \in X} p(x) \log_2 p(x) \]

\[ = H_2(X) \]

(the two summations above are essentially the same as this single summation).
(b) For equality to hold, i.e. $H_2(Y) = H_2(X)$, we need equality in 2.

This happens when the summation on $x \in X : g(x) = y$ involves only one term for all $y \in Y$, i.e. $\forall y$, there exists a unique $x \in X$ such that $g(x) = y$.

In other words, the function $g : X \to Y$ is one-to-one!