\[ M_x = \begin{bmatrix} \Pr(x_1) \\ \Pr(x_2) \\ \vdots \\ \Pr(x_n) \end{bmatrix} \quad \text{Input distribution} \]

\[ M_y = \begin{bmatrix} \Pr(y_1) \\ \Pr(y_2) \\ \vdots \\ \Pr(y_m) \end{bmatrix} \quad \text{Output distribution} \]

\[ M_y = \begin{bmatrix} \Pr(y_1|x_1) & \Pr(y_2|x_1) & \cdots & \Pr(y_m|x_1) \\ \vdots & \vdots & \ddots & \vdots \\ \Pr(y_1|x_n) & \Pr(y_2|x_n) & \cdots & \Pr(y_m|x_n) \end{bmatrix} \quad \text{PMF of the channel output} \]

\[ M_y = M_c \times M_x \quad \text{PMF of the channel input} \]

\[ P_{ij} = \Pr[y_i = y_j, x = x_j] \quad \text{Input to output PMF} \]

Example: Binary symmetric channel

\[ X = \{0, 1\} \]
\[ Y = \{0, 1\} \]
\[ N = R = 2 \]

Channel matrix:

\[ M_c = \begin{bmatrix} p & 1-p \\ 1-p & p \end{bmatrix} \]

\[ e = 1 - p \]
Michael Information:

The entropy of a source \( X \) with alphabet \( \{x_1, x_2, \ldots, x_M\} \) is

\[
H(X) = - \sum_{i=1}^{M} \pi_i(x_i) \log_2(\pi_i(x_i))
\]

- Average uncertainty per symbol of the source

Suppose that the receiver (at the channel output) receives symbol \( y_i \): The reception of \( y_i \) has changed the PMF of the source.

Before \( y_i \) is received:

\[
\begin{align*}
\pi_1(x_1) & \quad \pi_2(x_2) \\
\pi_3(x_3) & \quad \pi_4(x_4)
\end{align*}
\]

- Prior probabilities

Example:

\[\pi_i(x_i) = \begin{cases} 0.5 & \text{if } x_i = 1 \\ 0.5 & \text{if } x_i = 0 \end{cases}\]

<table>
<thead>
<tr>
<th>( X )</th>
<th>0.33</th>
<th>1.00</th>
</tr>
</thead>
<tbody>
<tr>
<td>( Y = 1 )</td>
<td>0.01</td>
<td>0.01</td>
</tr>
</tbody>
</table>

After \( y_i \) is received:

\[
\begin{align*}
\pi_1(x_1) | Y = y_i & = \frac{\pi(x_1, Y = y_i)}{\pi(Y = y_i)} \\
\pi_2(x_2) | Y = y_i & = \frac{\pi(x_2, Y = y_i)}{\pi(Y = y_i)} \\
\pi_3(x_3) | Y = y_i & = \frac{\pi(x_3, Y = y_i)}{\pi(Y = y_i)} \\
\pi_4(x_4) | Y = y_i & = \frac{\pi(x_4, Y = y_i)}{\pi(Y = y_i)}
\end{align*}
\]

After \( y_i \) is received:

\[
\begin{align*}
\pi_1(x_1) & = \frac{\pi(x_1, Y = y_i)}{\pi(Y = y_i)} = \frac{0.9}{0.9} = 1.0 \\
\pi_2(x_2) & = \frac{\pi(x_2, Y = y_i)}{\pi(Y = y_i)} = \frac{0.01}{0.9} = 0.01 \\
\pi_3(x_3) & = \frac{\pi(x_3, Y = y_i)}{\pi(Y = y_i)} = \frac{0.01}{0.9} = 0.01 \\
\pi_4(x_4) & = \frac{\pi(x_4, Y = y_i)}{\pi(Y = y_i)} = \frac{0.01}{0.9} = 0.01
\end{align*}
\]

\[
\begin{align*}
\pi(x_i) | Y = y_i & = \frac{\pi(x_i, Y = y_i)}{\pi(Y = y_i)} \\
\pi(x_1) | Y = y_i & = \frac{0.9}{0.9} = 1.0 \\
\pi(x_2) | Y = y_i & = \frac{0.01}{0.9} = 0.01 \\
\pi(x_3) | Y = y_i & = \frac{0.01}{0.9} = 0.01 \\
\pi(x_4) | Y = y_i & = \frac{0.01}{0.9} = 0.01
\end{align*}
\]

- Posterior probabilities

Example:

\[
\begin{align*}
\pi(x = 1 | Y = 1) & = \frac{\pi(x = 1, Y = 1)}{\pi(Y = 1)} = \frac{0.9}{0.9} = 1.0 \\
\pi(x = 0 | Y = 1) & = \pi(x = 0) \cdot \pi(Y = 1 | x = 0) + \pi(x = 0) \cdot \pi(Y = 1 | x = 0) \\
& = 0.5 \times 0.33 + 0.5 \times 0.01 = 0.5 \\
\pi(x = 0 | Y = 1) & = \frac{\pi(x = 1 | Y = 0) \cdot \pi(x = 0)}{\pi(Y = 1)} = \frac{0.01 \times 0.5}{0.01} = 0.01
\end{align*}
\]
After the symbol $y_i$ is received, the entropy of the source $x$ becomes:

\[ H(x | y_i) = - \sum_{j=1}^{M} p_{y_i}(x_j | y_i) \log_2 \left( p_{y_i}(x_j | y_i) \right) \]  

Expression (*) represents the entropy of the source after the reception of a single symbol ($y_i$). Of greater practical interest is the average entropy across all received symbols:

\[ H(x | y) = \frac{1}{R} \sum_{i=1}^{R} H(x | y_i) \]

\[ = - \sum_{i=1}^{R} \sum_{j=1}^{M} \frac{p_{y_i}(y_i) \cdot p_{y_i}(x_j | y_i)}{p_{y_i}(y_i)} \log_2 \left( \frac{p_{y_i}(x_j | y_i)}{p_{y_i}(x_j, y_i)} \right) \]

\[ p(x_j | y_i) = \frac{p(x_j, y_i)}{p(y_i)} \]

\[ H(x) - H(x | y) \leq \text{information gain} \]

**Example:** Consider a binary symmetric channel

\[ M = \begin{bmatrix} p & 1-p \\ 1-p & p \end{bmatrix} \]

\[ p : \text{probability of correct reception} \]

\[ \alpha = 2, \quad \beta = 0.5 \]
Assuming $\sum_i X_i \in \{0, 1\}$

$$
H(X) = \text{1 bit} = -0.5 \cdot \log_2(0.5) - 0.5 \log_2(0.5)
= \frac{1}{2} + \frac{1}{2} = 1
$$

Conditional entropy

$$
H(X|Y) = \sum_{i=1}^{2} \Pr(Y_i) \cdot H(X|Y_i)
$$

**Step 1:** Find the output probabilities

$$
\Pr(Y = 0) = \Pr(Y = 0 | X = 0) \cdot \Pr(X = 0) + \Pr(Y = 0 | X = 1) \cdot \Pr(X = 1)
= \frac{1}{2} \cdot p + \frac{1}{2} \cdot (1 - p) = \frac{1}{2}
$$

$$
\Pr(Y = 1) = \frac{1}{2}
$$

**Step 2:**

$$
H(X|Y = 0) = - \Pr(X = 0 | Y = 0) \cdot \log_2(\Pr(X = 0 | Y = 0))
- \Pr(X = 1 | Y = 0) \cdot \log_2(\Pr(X = 1 | Y = 0))
$$

Find $\Pr(X = 0 | Y = 0) = \frac{\Pr(X = 0, Y = 0)}{\Pr(Y = 0)} = \frac{\Pr(Y = 0 | X = 0) \cdot \Pr(X = 0)}{\Pr(Y = 0)} = \frac{1}{2} \cdot \Pr(Y = 0 | X = 0) \cdot \Pr(X = 0)
= P$

$\Pr(X = 1 | Y = 0) = 1 - P$

$$
H(X|Y = 0) = - P \log_2(P) - (1 - P) \log_2(1 - P)
$$

Also,

$$
H(X|Y = 1) = - P \log_2(P) - (1 - P) \log_2(1 - P)
$$

**Step 3:**

$$
H(X|Y) = \frac{1}{2} \cdot H(X|Y = 0) + \frac{1}{2} \cdot H(X|Y = 1)
$$
\[ H(X|Y) = -p \log_2(p) - (1-p) \log_2(1-p) \]

2) \( p = 1 \Rightarrow H(X|Y) = 0 \) → Uncertainty about what was sent is zero.
   (Channel is ideal)

   * When we receive the symbol, we can conclude with absolute certainty what has been sent.

2) \( p = 0 \Rightarrow H(X|Y) = 0 \) : Uncertainty is zero.

3) \( p = \frac{1}{2} \Rightarrow H(X|Y) = 1 = H(X) \)

   The receiver of \( Y \) does not change the uncertainty about the source. By receiving \( Y \), the receiver does not gain any new information about the input. → Worst case scenario

**Mutual Information Definition:**

\[ I(X,Y) = H(X) - H(X|Y) \]

\( H(X) \) is Entropy of source

\( H(X|Y) \) is Unresolved entropy of the source
For ideal channels

\[ I(X;Y) = H(X) \]

⇒ Channel has delivered all info from source to the receiver.

For all other channels:

\[ I(X;Y) < H(X) \]

There is a loss of information due to channel distortion.

Substituting expressions for \( H(X) \) and \( H(X|Y) \) one obtains:

\[ I(X;Y) = -\sum_{j=1}^{M} \left( p(x_j) \log_2(p(x_j)) \right) \]

\[ \sum_{i=1}^{L} \sum_{j=1}^{M} p(x_i,y_j) \log_2 \left( \frac{p(x_i|y_j)}{p(x_i)} \right) \]

Aside: \( p(x_j) = \sum_{i=1}^{R} p(x_j|y_i) \)

\[ p(x_j|y_i) = \frac{\sum_{i=1}^{R} p(x_j|y_i) p(y_i)}{p(x_j)} \]

\((***)\) is the most common way of defining mutual information.

Can be estimated using measurement of the channel.
Source \( \mathbf{X} \)

\[ \begin{align*}
\text{Output:} & \quad Y_1, Y_2, \ldots \\
\text{Input:} & \quad X_1, X_2, \ldots \\
\text{Repeat experiment for N times:} & \\
\Pr(X_1) & \approx \frac{N_{X_1}}{N} \\
\Pr(X_2) & \approx \frac{N_{X_2}}{N} \\
\vdots & \\
\Pr(Y_1) & \approx \frac{N_{Y_1}}{N} \\
\Pr(Y_2) & \approx \frac{N_{Y_2}}{N} \\
\vdots & \\
\text{Probabilities for pairs } (X_j, Y_i) \rightarrow \Pr(X_j, Y_i) \approx \frac{N_{(X_j, Y_i)}}{N}
\end{align*} \]

Example: Consider a binary channel (is not symmetric)

\[ \Pr(X = 0) = \Pr(Y = 1) = \frac{1}{2} \]

At the channel output, we observe the following probabilities:

\[ \Pr(Y = 0) = \Pr(Y = 1) = \frac{1}{2} \]

\[
\begin{bmatrix}
\Pr(X_j, Y_i) \\
1 \leq j \leq 2 \\
2 \leq i \leq 2
\end{bmatrix}
\begin{bmatrix}
Y_1 \\
0.4 \\
Y_2 \\
0.2 \\
X_2 \\
0.2 \\
X_1 \\
0.2
\end{bmatrix}
\]

\[
\begin{bmatrix}
Y_1 \\
0.4 \\
Y_2 \\
0.2 \\
X_2 \\
0.2 \\
X_1 \\
0.2
\end{bmatrix}
\begin{bmatrix}
0.2 \\
0.4 \\
0.2 \\
0.2 \\
0.4 \\
0.2 \\
0.2 \\
0.2
\end{bmatrix}
\]

\[
\begin{bmatrix}
Y_1 \\
0.4 \\
Y_2 \\
0.2 \\
X_2 \\
0.2 \\
X_1 \\
0.2
\end{bmatrix}
\begin{bmatrix}
0.2 \\
0.4 \\
0.2 \\
0.2 \\
0.4 \\
0.2 \\
0.2 \\
0.2
\end{bmatrix}
\]

\[
I(X; Y) = \sum_{j=1}^{2} \sum_{i=1}^{2} \Pr(X_j, Y_i) \cdot \left[ \log_2 \left( \frac{\Pr(X_j, Y_i)}{\Pr(X_j) \cdot \Pr(Y_i)} \right) \right]
\]

\[
\Pr(X_j) = \Pr(Y_i) = \frac{1}{2}, \quad X_j, Y_i \in \{1, 2\}
\]

\[
\log_2 \left( \frac{\Pr(X_j) \cdot \Pr(Y_i)}{\Pr(X_j) \cdot \Pr(Y_i)} \right) = -2
\]
\[ I(x; X) = 0.4 \times \left[ \log_2(0.4) + 2 \right] + 0.2 \times \left[ \log_2(0.2) + 2 \right] \times 3 = 0.078 \] \Rightarrow \text{very small}

→ Here channel is pretty broken this is why \( I(x; Y) \) small

\[
\Pr (x_i, y_i) = \begin{bmatrix}
(0, 0) & 0.1 \\
(1, 0) & 0.5 \\
(1, 1) & 0.4
\end{bmatrix} \Rightarrow I(x; y) = 1 \text{ bit}
\]

\( \neq M_{\text{channel}} \) not \( \text{Ideal Channel} \)

\( p(y_i | x_j) \)