4F5: Advanced Communications and Coding
Handout 3: Discrete Channels, Channel Capacity

Ramji Venkataramanan

Signal Processing and Communications Lab
Department of Engineering
ramji.v@eng.cam.ac.uk

Michaelmas Term 2014
End-to-End Communication System

Compressor

Channel

Decompressor

Transmitter

Receiver

Compression

Transmission
Transmitter does two things:

1. **Coding**: Adding redundancy to the data bits to protect against noise

2. **Modulation**: Transforming the coded bits into waveforms. E.g. PSK, PAM, QAM etc. are modulation schemes (we’ll study these later)
Optimally, we should design the coding and the modulation in a combined manner.

However, for engineering simplicity, the modulation scheme and the error-correcting code are often chosen separately.
Because of noise in the channel, some bits at the output of the demodulator are in error.

Every modulation scheme has an associated probability of error, say $p$, that we can estimate theoretically or empirically.
Because of noise in the channel, some bits at the output of the demodulator are in error.

Every modulation scheme has an associated *probability of error*, say $p$, that we can estimate theoretically or empirically.
Binary Symmetric Channel (BSC)

Thus a very important practical channel is the following:

\[
\begin{align*}
X = 0 & \quad \rightarrow \quad Y = 0 & & (1 - p) \\
X = 1 & \quad \rightarrow \quad Y = 1 & & (1 - p) \\
Y = 0 & \quad \rightarrow \quad X = 0 & & p \\
Y = 1 & \quad \rightarrow \quad X = 1 & & p
\end{align*}
\]

\[
P(Y = 0|X = 0) = 1 - p, \quad P(Y = 1|X = 1) = 1 - p
\]

\[
P(Y = 1|X = 0) = p, \quad P(Y = 0|X = 1) = p
\]

- \(p\) is the “crossover probability”; the channel is called BSC(\(p\))
- KEY Question: How to design good error-correcting codes for the BSC?
Example: Repetition Code

\[ X = 0 \quad \xrightarrow{0.9} \quad Y = 0 \]
\[ X = 1 \quad \xrightarrow{0.9} \quad Y = 1 \]

Data: 0 1 1 0 0 . . .
Coded bits: 000 111 111 000 000 . . .
Received bits: 001 101 111 011 000 . . .
Decoded bits: 0 1 1 1 0 . . .

Data rate = \frac{1}{3} \text{ bits/transmission}

Probability of bit decoding error:
\[
\left( \frac{3}{2} \right)^2 \left( \frac{1}{9} \right) + \left( \frac{3}{3} \right)^3 \left( \frac{1}{9} \right)^3 = 0.028
\]
Example: Repetition Code

\[ X = 0 \quad \Rightarrow \quad Y = 0 \]
\[ X = 1 \quad \Rightarrow \quad Y = 1 \]

(1, 3) Repetition Code

<table>
<thead>
<tr>
<th>Data</th>
<th>0</th>
<th>1</th>
<th>1</th>
<th>0</th>
<th>0...</th>
</tr>
</thead>
<tbody>
<tr>
<td>Coded bits</td>
<td>000</td>
<td>111</td>
<td>111</td>
<td>000</td>
<td>000...</td>
</tr>
<tr>
<td>Received bits</td>
<td>001</td>
<td>101</td>
<td>111</td>
<td>011</td>
<td>000...</td>
</tr>
</tbody>
</table>

- Data bit wrongly decoded if channel flips two or more bits in its “codeword”
Example: Repetition Code

\[ X = 0 \quad \text{\rightarrow} \quad Y = 0 \]
\[ X = 1 \quad \text{\rightarrow} \quad Y = 1 \]

(1, 3) Repetition Code

Data: 0 1 1 0 0...
Coded bits: 000 111 111 000 000...
Received bits: 001 101 111 011 000...
Decoded bits: 0 1 1 1 0...

Data bit wrongly decoded if channel flips two or more bits in its “codeword”
Example: Repetition Code

\[
\begin{align*}
X = 0 & \quad \text{.9} \quad Y = 0 \\
X = 1 & \quad \text{.9} \quad Y = 1
\end{align*}
\]

\[(1, 3)\] Repetition Code

<table>
<thead>
<tr>
<th>Data</th>
<th>0</th>
<th>1</th>
<th>1</th>
<th>0</th>
<th>0…</th>
</tr>
</thead>
<tbody>
<tr>
<td>Coded bits</td>
<td>000</td>
<td>111</td>
<td>111</td>
<td>000</td>
<td>000…</td>
</tr>
<tr>
<td>Received bits</td>
<td>001</td>
<td>101</td>
<td>111</td>
<td>011</td>
<td>000…</td>
</tr>
<tr>
<td>Decoded bits</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0…</td>
</tr>
</tbody>
</table>

- Data bit wrongly decoded if channel flips two or more bits in its “codeword”
- Probability of bit decoding error: \( \left( \frac{3}{2} \right) (.1)^2 (.9) + \left( \frac{3}{3} \right) (.1)^3 = 0.028 \)
- Data rate = \( \frac{1}{3} \) bits/transmission
If we use a \((1,9)\) repetition code:

Probability of bit error = 0.0009; Data rate = \(\frac{1}{9}\) bits/transmission
If we use a \((1, 9)\) repetition code:

Probability of bit error = 0.0009; Data rate = \(\frac{1}{9}\) bits/transmission

As we increase repetition code length \(n\):

- Probability of error goes to 0 😊
- Data rate = \(\frac{1}{n}\), which also goes to 0 😒
If we use a $(1, 9)$ repetition code:
Probability of bit error $= 0.0009$; Data rate $= \frac{1}{9}$ bits/transmission

As we increase repetition code length $n$:
- Probability of error goes to 0 ☹
- Data rate $= \frac{1}{n}$, which also goes to 0 ☹

Can we have $P(\text{error}) \rightarrow 0$ at strictly positive data rate?
Yes! [Shannon ’48]

For the BSC(0.1) above, we can communicate at rate of as high as 0.53 bit/transmission with \textit{arbitrarily} small probability of error.
A discrete memoryless channel (DMC) is a system consisting of an input alphabet $\mathcal{X}$, output alphabet $\mathcal{Y}$, and a set of transition probabilities

$$P_{Y|X}(b|a) = Pr(Y = b|X = a) \text{ for all } a \in \mathcal{X} \text{ and } b \in \mathcal{Y}.$$
A discrete memoryless channel (DMC) is a system consisting of an input alphabet $\mathcal{X}$, output alphabet $\mathcal{Y}$, and a set of transition probabilities

$$P_{Y|X}(b|a) = Pr(Y = b|X = a) \text{ for all } a \in \mathcal{X} \text{ and } b \in \mathcal{Y}.$$ 

Memoryless means that the channel acts independently on each input. For each time instant $k = 1, 2, \ldots$

$$Pr(Y_k = y|X_k = x, X_{k-1}, \ldots, X_1, Y_{k-1}, \ldots, Y_1) = P_{Y|X}(y|x).$$

"Given all the past, the current output depends only on the current input"
DMC Examples

Binary Symmetric Channel (BSC):

A DMC can be described by a transition probability matrix, e.g.

\[
\begin{array}{cc}
0 & 0 \\
\hline
0 & 1 \\
\hline
1 & 0 \\
\end{array}
\]

\[
\begin{array}{cc}
p & 1-p \\
\hline
p & 1-p \\
\end{array}
\]

\[
\begin{array}{ccc}
P_{Y|X} & 0 & 1 \\
\hline
X & 0 & 1-p & p \\
1 & p & 1-p \\
\end{array}
\]
DMC Examples

Binary erasure channel (BEC):

\[
\begin{array}{c c c c}
0 & 1-e & 0 \\
1 & e & ?
\end{array}
\]

\[
\begin{array}{c c c c}
0 & 1-e & e & 0 \\
1 & 0 & e & 1-e
\end{array}
\]

- Like the BSC, the BEC is also practically important, e.g:
  - Erasure can model packet loss in networks
  - When the demodulator thinks the (real-valued) output symbol is too noisy, it can declare an erasure
Noisy Keyboard Channel: A useful toy example

Input alphabet with 26 symbols: $a, b, \ldots, z$

Each channel input is either received unchanged or transformed into next symbol:

$$P(Y = a | X = a) = P(Y = b | X = a) = \frac{1}{2},$$

$$P(Y = b | X = b) = P(Y = c | X = b) = \frac{1}{2},$$

\[ \vdots \]

$$P(Y = z | X = z) = P(Y = a | X = z) = \frac{1}{2}. $$
How can we communicate error-free over this channel?

13 / 18
How can we communicate error-free over this channel?

- A simple code: use only symbols $a, c, e, \ldots, y$
How can we communicate error-free over this channel?

- A simple code: use only symbols $a, c, e, \ldots, y$
- Can *noiselessly* convey one of 13 symbols per transmission
  \[ \Rightarrow \text{Transmission rate} = \log 13 \text{ bits/transmission} \]
- We will see that this is, in fact, the maximum possible rate
How can we communicate error-free over this channel?

- A simple code: use only symbols $a, c, e, \ldots, y$
- Can noiselessly convey one of 13 symbols per transmission
  \[ \Rightarrow \text{Transmission rate} = \log 13 \text{ bits/transmission} \]
- We will see that this is, in fact, the maximum possible rate

For a general DMC:

- We’ll construct a set of input sequences which have non-intersecting sets of output sequences with high prob.
- These input sequences are “non-confusable” analogous to inputs $a, c, e, \ldots$ in the keyboard channel.
The \textit{channel capacity} of a discrete memoryless channel is defined as

$$C = \max_{P_X} I(X; Y)$$

- $P_{Y|X}$ is defined by the DMC. Each input distribution $P_X$ yields a different joint distribution on $(X, Y)$ given by $P_X P_{Y|X}$. 
The channel capacity of a discrete memoryless channel is defined as

\[ C = \max_{P_X} I(X; Y) \]

- \( P_{Y|X} \) is defined by the DMC. Each input distribution \( P_X \) yields a different joint distribution on \((X, Y)\) given by \( P_X P_{Y|X} \).
- For now, \( C \) is just a mathematical quantity, defined for each DMC.
- We will show that \( C \) is the maximum transmission rate over the DMC if you want arbitrarily small probability of error.

First, let’s compute \( C \) for some examples . . .
Example 1: Noiseless Binary Channel

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

What \( P_X \) maximises \( H(X) \)?

Ans: 
\( P_X = \left( \frac{1}{2}, \frac{1}{2} \right) \)

Therefore 
\[ C = \max P_X I(X; Y) = \max P_X H(X) = 1 \text{ bit/transmission} \]
Example 1: Noiseless Binary Channel

\[ I(X; Y) = H(X) - H(X|Y) = H(X) \quad \text{(why is } H(X|Y) = 0?) \]

- What \( P_X \) maximises \( H(X) \)?
Example 1: Noiseless Binary Channel

\[ I(X; Y) = H(X) - H(X|Y) = H(X) \]  (why is \( H(X|Y) = 0 \)?)

- What \( P_X \) maximises \( H(X) \)?  
  Ans: \( P_X = \left( \frac{1}{2}, \frac{1}{2} \right) \)
- Therefore

\[ C = \max_{P_X} I(X; Y) = \max_{P_X} H(X) = 1 \text{ bit/transmission}. \]
Example 2: BSC

\[ I(X; Y) = H(Y) - H(Y|X) \]
\[ = H(Y) - \sum_{x \in \{0,1\}} P_X(x) H(Y|X = x) \]
Example 2: BSC

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

\[
= H(Y) - \sum_{x \in \{0, 1\}} P_X(x)H(Y|X = x)
\]

\[
\text{why?} \quad \overset{=} \Rightarrow H(Y) - \sum_{x \in \{0, 1\}} P_X(x)H_2(p)
\]
Example 2: BSC

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

\[ = H(Y) - \sum_{x \in \{0,1\}} P_X(x)H(Y|X = x) \]

\[ = H(Y) - \sum_{x \in \{0,1\}} P_X(x)H_2(p) \]

\[ = H(Y) - H_2(p) \]
Example 2: BSC

\[ I(X; Y) = H(Y) - H(Y|X) \]
\[ = H(Y) - \sum_{x \in \{0,1\}} P_X(x)H(Y|X = x) \]

\[ \text{why?} \quad \equiv H(Y) - \sum_{x \in \{0,1\}} P_X(x)H_2(p) \]
\[ = H(Y) - H_2(p) \]
\[ \leq 1 - H_2(p). \]

- Maximum attained when \( P_X = (\frac{1}{2}, \frac{1}{2}) \) \( \Rightarrow \ C = 1 - H_2(p) \)
- For \( p = 0.1 \), \( C = 0.531 \) bits/transmission
Example 3: Noisy Keyboard Channel

$I(X; Y) = H(Y) - H(Y|X)$

What $P_X$ maximises $H(Y)$?

We have two choices!

1) $P_X = (1/26, 1/26, \ldots, 1/26)$
2) $P_X = (1/13, 0, \ldots, 1/13, 0)$ [We used this one earlier]

Both yield $P_Y = (1/26, 1/26, \ldots, 1/26)$

$C = \max P_X I(X; Y) = \log(26) - 1 = \log 13$ bits/transmission.

MORAL: Maximising input distribution may not be unique!
Example 3: Noisy Keyboard Channel

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

\[
= H(Y) - 1 \leq \log(26) - 1
\]

What \( P_X \) maximises \( H(Y) \)?
Example 3: Noisy Keyboard Channel

\[ I(X; Y) = H(Y) - H(Y|X) \]
\[ = H(Y) - 1 \leq \log(26) - 1 \]

- What \( P_X \) maximises \( H(Y) \)? We have two choices!
  1) \( P_X = \left( \frac{1}{26}, \frac{1}{26}, \ldots, \frac{1}{26} \right) \)
  2) \( P_X = \left( \frac{1}{13}, 0, \ldots, \frac{1}{13}, 0 \right) \) [We used this one earlier]

- Both yield \( P_Y = \left( \frac{1}{26}, \frac{1}{26}, \ldots, \frac{1}{26} \right) \)

\[ C = \max_{P_X} I(X; Y) = \log(26) - 1 = \log 13 \text{ bits/transmission.} \]

MORAL: Maximising input distribution may not be unique!
You can now do Questions 1–13 in Examples Paper 1.