COMP2610/6261 - Information Theory

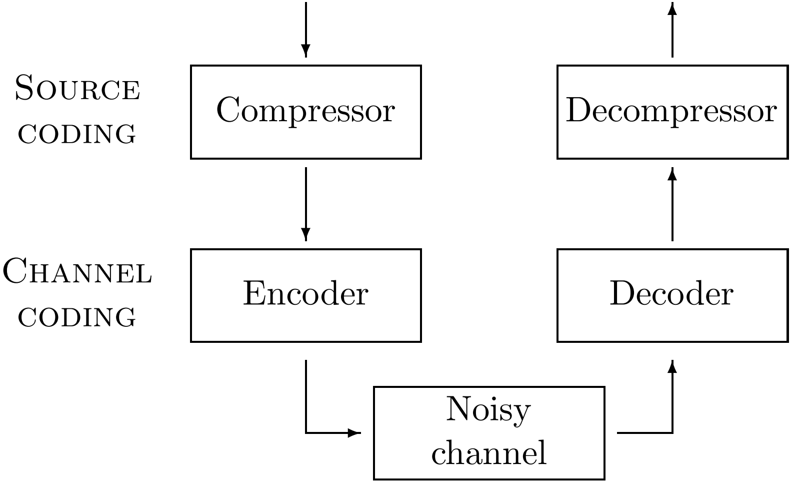
Lecture 18: Noisy Channels

Mark Reid and Aditya Menon

Research School of Computer Science
The Australian National University

October 1st, 2014

The Big Picture



Channels

A **discrete channel** Q consists of an *input alphabet* $\mathcal{X} = \{a_1, \ldots, a_I\}$, an *output alphabet* $\mathcal{Y} = \{b_1, \ldots, b_J\}$ and *transistion probabilities* $P(y|x)$. The channel Q can be expressed as a matrix

$$Q_{j,i} = P(y = b_j | x = a_i)$$

Channels

A **discrete channel** Q consists of an *input alphabet* $\mathcal{X} = \{a_1, \ldots, a_I\}$, an *output alphabet* $\mathcal{Y} = \{b_1, \ldots, b_J\}$ and *transistion probabilities* $P(y|x)$. The channel Q can be expressed as a matrix

$$Q_{j,i} = P(y = b_j | x = a_i)$$

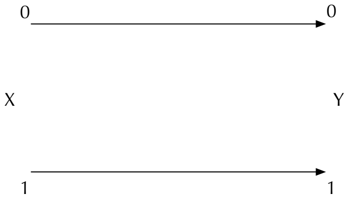
Example: A channel Q with inputs $\mathcal{X} = \{a_1, a_2, a_3\}$, outputs $\mathcal{Y} = \{b_1, b_2\}$, and transition probabilities expressed by the matrix

$$Q = \begin{bmatrix} 0.8 & 0.5 & 0.2 \\ 0.2 & 0.5 & 0.8 \end{bmatrix}$$

So $P(b_1|a_1) = 0.8 = P(b_2|a_3)$ and $P(b_1|a_2) = P(b_2|a_2) = 0.5$.

The Binary Noiseless Channel

One of the simplest channels is the **Binary Noiseless Channel**. The received symbol is always equal to the transmitted symbol – there is no probability of error, hence *noiseless*.

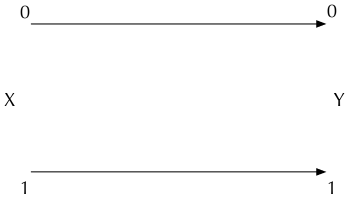


Inputs $\mathcal{X} = \{0, 1\}$; Outputs $\mathcal{Y} = \{0, 1\}$;
Transition probabilities

$$Q = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

The Binary Noiseless Channel

One of the simplest channels is the **Binary Noiseless Channel**. The received symbol is always equal to the transmitted symbol – there is no probability of error, hence *noiseless*.



Inputs $\mathcal{X} = \{0, 1\}$; Outputs $\mathcal{Y} = \{0, 1\}$; Transition probabilities

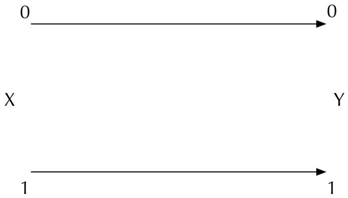
$$Q = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

What was transmitted over the channel if 0000 1111 was received?

$$\xrightarrow{Q} 0000\,1111$$

The Binary Noiseless Channel

One of the simplest channels is the **Binary Noiseless Channel**. The received symbol is always equal to the transmitted symbol – there is no probability of error, hence *noiseless*.



Inputs $\mathcal{X} = \{0, 1\}$; Outputs $\mathcal{Y} = \{0, 1\}$;
Transition probabilities

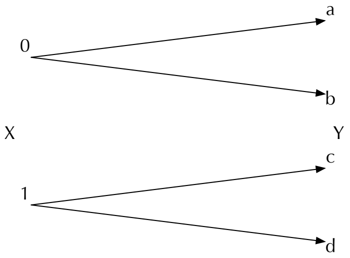
$$Q = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

What was transmitted over the channel if 0000 1111 was received?

$$0000\,1111 \xrightarrow{Q} 0000\,1111$$

The Noisy Non-overlapping Channel

Even if there is some uncertainty about the output given the input, it may still be possible to perfectly infer what was transmitted.

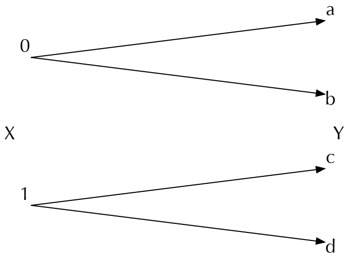


Inputs $\mathcal{X} = \{0, 1\}$; Outputs $\mathcal{Y} = \{\mathtt{a}, \mathtt{b}, \mathtt{c}, \mathtt{d}\}$; Transition probabilities

$$Q = \begin{bmatrix} 0.5 & 0 \\ 0.5 & 0 \\ 0 & 0.5 \\ 0 & 0.5 \end{bmatrix}$$

The Noisy Non-overlapping Channel

Even if there is some uncertainty about the output given the input, it may still be possible to perfectly infer what was transmitted.



Inputs $\mathcal{X} = \{0, 1\}$; Outputs $\mathcal{Y} = \{\mathtt{a}, \mathtt{b}, \mathtt{c}, \mathtt{d}\}$; Transition probabilities

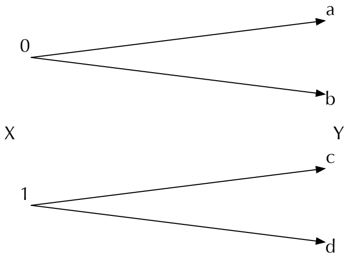
$$Q = \begin{bmatrix} 0.5 & 0 \\ 0.5 & 0 \\ 0 & 0.5 \\ 0 & 0.5 \end{bmatrix}$$

What was transmitted over the channel if `abba ddcd` was received?

$$\xrightarrow{Q} \mathtt{abba\,ddcd}$$

The Noisy Non-overlapping Channel

Even if there is some uncertainty about the output given the input, it may still be possible to perfectly infer what was transmitted.



Inputs $\mathcal{X} = \{0, 1\}$; Outputs $\mathcal{Y} = \{\mathtt{a}, \mathtt{b}, \mathtt{c}, \mathtt{d}\}$; Transition probabilities

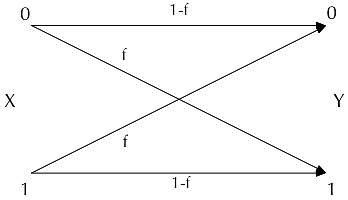
$$Q = \begin{bmatrix} 0.5 & 0 \\ 0.5 & 0 \\ 0 & 0.5 \\ 0 & 0.5 \end{bmatrix}$$

What was transmitted over the channel if abba ddcd was received?

$$0000\,1111 \xrightarrow{Q} \mathtt{abba\,ddcd}$$

The Binary Symmetric Channel

Each symbol sent across a **binary symmetric channel** has a chance of being "flipped" to its counterpart ($0 \to 1$; $1 \to 0$)

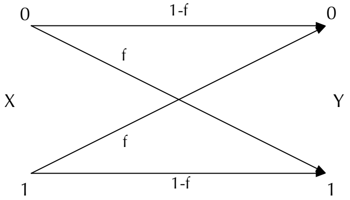


Inputs $\mathcal{X} = \{0, 1\}$; Outputs $\mathcal{Y} = \{0, 1\}$;
Transition probabilities with $P(\text{flip}) = f$

$$Q = \begin{bmatrix} 1 - f & f \\ f & 1 - f \end{bmatrix}$$

The Binary Symmetric Channel

Each symbol sent across a **binary symmetric channel** has a chance of being "flipped" to its counterpart ($0 \to 1$; $1 \to 0$)



Inputs $\mathcal{X} = \{0, 1\}$; Outputs $\mathcal{Y} = \{0, 1\}$;
Transition probabilities with $P(\text{flip}) = f$

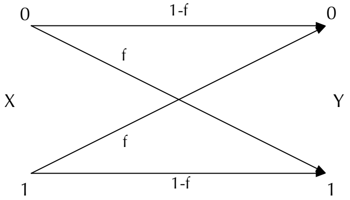
$$Q = \begin{bmatrix} 1-f & f \\ f & 1-f \end{bmatrix}$$

What was most likely transmitted over the channel if 0010 1001 was received, assuming $f = 0.1$ and $P(x = 0) = P(x = 1) = 0.5$?

$$\xrightarrow{Q} 0010\,1001$$

The Binary Symmetric Channel

Each symbol sent across a **binary symmetric channel** has a chance of being "flipped" to its counterpart ($0 \to 1$; $1 \to 0$)



Inputs $\mathcal{X} = \{0, 1\}$; Outputs $\mathcal{Y} = \{0, 1\}$;
Transition probabilities with $P(\text{flip}) = f$

$$Q = \begin{bmatrix} 1-f & f \\ f & 1-f \end{bmatrix}$$

What was most likely transmitted over the channel if 0010 1001 was received, assuming $f = 0.1$ and $P(x = 0) = P(x = 1) = 0.5$?

$$0000\,1111 \xrightarrow{Q} 0010\,1001$$

Inferring the Input

Suppose we know the $P(x)$ — the probability x is transmitted.

Given a particular output $y \in \mathcal{Y}$ received over a channel Q, how likely was it that input $x \in \mathcal{X}$ was transmitted?

Inferring the Input

Suppose we know the $P(x)$ — the probability x is transmitted.
Given a particular output $y \in \mathcal{Y}$ received over a channel Q, how likely was it that input $x \in \mathcal{X}$ was transmitted?

$$P(x|y) = \frac{P(y|x)P(x)}{P(y)} = \frac{P(y|x)P(x)}{\sum_{x' \in \mathcal{X}} P(y|x')P(x')}$$

Inferring the Input

Suppose we know the $P(x)$ — the probability x is transmitted.
Given a particular output $y \in \mathcal{Y}$ received over a channel Q, how likely was it that input $x \in \mathcal{X}$ was transmitted?

$$P(x|y) = \frac{P(y|x)P(x)}{P(y)} = \frac{P(y|x)P(x)}{\sum_{x' \in \mathcal{X}} P(y|x')P(x')}$$

Example: Suppose $P(x = 0) = P(x = 1) = 0.5$. What are the probability that a $x = 0$ was transmitted over a *binary symmetric channel* Q with $f = 0.1$ given that a $y = 0$ was received?

Inferring the Input

Suppose we know the $P(x)$ — the probability x is transmitted.
Given a particular output $y \in \mathcal{Y}$ received over a channel Q, how likely was it that input $x \in \mathcal{X}$ was transmitted?

$$P(x|y) = \frac{P(y|x)P(x)}{P(y)} = \frac{P(y|x)P(x)}{\sum_{x' \in \mathcal{X}} P(y|x')P(x')}$$

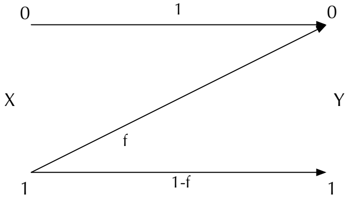
Example: Suppose $P(x = 0) = P(x = 1) = 0.5$. What are the probability that a $x = 0$ was transmitted over a *binary symmetric channel* Q with $f = 0.1$ given that a $y = 0$ was received?

$$P(x = 0|y = 0) = \frac{0.9 \times 0.5}{0.9 \times 0.5 + 0.1 \times 0.5} = 0.9$$

Similarly, $P(x = 1|y = 1) = 0.9$.

The Z Channel

Symbols may be corrupted over the channel asymmetrically.

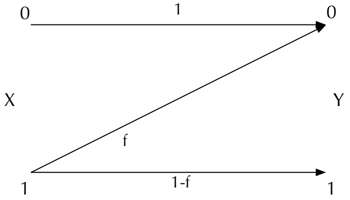


Inputs $\mathcal{X} = \{0, 1\}$; Outputs $\mathcal{Y} = \{0, 1\}$;
Transition probabilities

$$Q = \begin{bmatrix} 1 & f \\ 0 & 1 - f \end{bmatrix}$$

The Z Channel

Symbols may be corrupted over the channel asymmetrically.



Inputs $\mathcal{X} = \{0, 1\}$; Outputs $\mathcal{Y} = \{0, 1\}$;
Transition probabilities

$$Q = \begin{bmatrix} 1 & f \\ 0 & 1 - f \end{bmatrix}$$

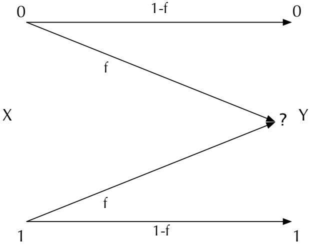
Inferring the input: Clearly $P(x = 1|y = 1) = 1$ but

$$P(x = 0|y = 0) = \frac{P(y = 0|x = 0)P(x = 0)}{\sum_{x' \in \mathcal{X}} P(y = 0|x')P(x')} = \frac{P(x = 0)}{P(x = 0) + f\, P(x = 1)}$$

So $P(x = 0|y = 0) \to 1$ as $f \to 0$, and goes to $P(x = 0)$ as $f \to 1$

The Binary Erasure Channel

We can model a channel which "erases" bits by letting one of the output symbols be the symbol '?' with associated probability f. The receiver knows which bits are erased.



Inputs $\mathcal{X} = \{0, 1\}$; Outputs $\mathcal{Y} = \{0, ?, 1\}$;
Transition probabilities

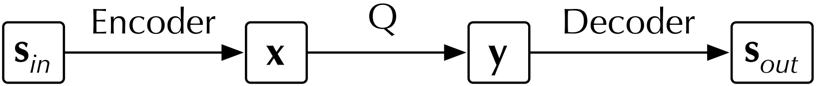
$$Q = \begin{bmatrix} 1-f & 0 \\ f & f \\ 0 & 1-f \end{bmatrix}$$

Example:

$$0000\,1111 \xrightarrow{Q} 00?0\,?11?$$

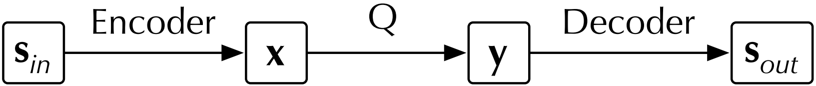
Communicating over Noisy Channels

Suppose we know we have to communicate over some channel Q and we want build an *encoder/decoder* pair to reliably send a message **s** over Q.



Communicating over Noisy Channels

Suppose we know we have to communicate over some channel Q and we want build an *encoder/decoder* pair to reliably send a message $\mathbf{s}$ over Q.

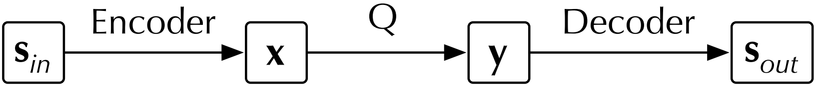


Reliability is measured via **probability of error** — that is, the probability of incorrectly decoding $\mathbf{s}_{out}$ given $\mathbf{s}_{in}$ as input:

$$P(\mathbf{s}_{out} \neq \mathbf{s}_{in}) = \sum_{\mathbf{s}} P(\mathbf{s}_{out} \neq \mathbf{s}_{in} | \mathbf{s}_{in}) P(\mathbf{s}_{in})$$

Communicating over Noisy Channels

Suppose we know we have to communicate over some channel Q and we want build an *encoder/decoder* pair to reliably send a message $\mathbf{s}$ over Q.



Reliability is measured via **probability of error** — that is, the probability of incorrectly decoding $\mathbf{s}_{out}$ given $\mathbf{s}_{in}$ as input:

$$P(\mathbf{s}_{out} \neq \mathbf{s}_{in}) = \sum_{\mathbf{s}} P(\mathbf{s}_{out} \neq \mathbf{s}_{in} | \mathbf{s}_{in}) P(\mathbf{s}_{in})$$

Example:

Let $\mathcal{S} = \{\mathtt{a}, \mathtt{b}\}$, with *encoder*: $\mathtt{a} \rightarrow 0$; $\mathtt{b} \rightarrow 1$, *decoder*: $0 \rightarrow \mathtt{a}$; $1 \rightarrow \mathtt{b}$.
For binary symmetric Q with $f = 0.1$ and $(p_{\mathtt{a}}, p_{\mathtt{b}}) = (0.5, 0.5)$

$$P(\mathbf{s}_{in} \neq \mathbf{s}_{out}) = P(y = 1 | x = 0)\, p_{\mathtt{a}} + P(y = 0 | x = 1)\, p_{\mathtt{b}} = f = 0.1$$

A Simple Coding Scheme

Suppose $\mathbf{s} \in \{\mathtt{a}, \mathtt{b}\}$ and we encode by $\mathtt{a} \rightarrow 000$ and $\mathtt{b} \rightarrow 111$.
To decode we count the number of 1s and 0s and set all bits to the majority count to determine $\mathbf{s}$

$$\underbrace{000, 001, 010, 100}_{A} \rightarrow \mathtt{a} \quad \text{and} \quad \underbrace{111, 110, 101, 011}_{B} \rightarrow \mathtt{b}$$

A Simple Coding Scheme

Suppose $\mathbf{s} \in \{\mathtt{a}, \mathtt{b}\}$ and we encode by $\mathtt{a} \to \mathtt{000}$ and $\mathtt{b} \to \mathtt{111}$.
To decode we count the number of 1s and 0s and set all bits to the majority count to determine $\mathbf{s}$

$$\underbrace{\mathtt{000}, \mathtt{001}, \mathtt{010}, \mathtt{100}}_{A} \to \mathtt{a} \quad \text{and} \quad \underbrace{\mathtt{111}, \mathtt{110}, \mathtt{101}, \mathtt{011}}_{B} \to \mathtt{b}$$

If the channel Q is binary symmetric with $f = 0.1$ again

$$\begin{aligned} P(\mathbf{s}_{in} \neq \mathbf{s}_{out}) &= P(\mathbf{y} \in B|000)\, p_{\mathtt{a}} + P(\mathbf{y} \in A|111)\, p_{\mathtt{b}} \\ &= [f^3 + 3f^2(1-f)]p_{\mathtt{a}} + [f^3 + 3f^2(1-f)]p_{\mathtt{b}} \\ &= f^3 + 3f^2(1-f) = 0.028 \end{aligned}$$

A Simple Coding Scheme

Suppose $\mathbf{s} \in \{\mathtt{a}, \mathtt{b}\}$ and we encode by $\mathtt{a} \rightarrow \mathtt{000}$ and $\mathtt{b} \rightarrow \mathtt{111}$.
To decode we count the number of 1s and 0s and set all bits to the majority count to determine $\mathbf{s}$

$$\underbrace{\mathtt{000}, \mathtt{001}, \mathtt{010}, \mathtt{100}}_{A} \rightarrow \mathtt{a} \quad \text{and} \quad \underbrace{\mathtt{111}, \mathtt{110}, \mathtt{101}, \mathtt{011}}_{B} \rightarrow \mathtt{b}$$

If the channel Q is binary symmetric with $f = 0.1$ again

$$\begin{aligned} P(\mathbf{s}_{in} \neq \mathbf{s}_{out}) &= P(\mathbf{y} \in B | 000)\, p_{\mathtt{a}} + P(\mathbf{y} \in A | 111)\, p_{\mathtt{b}} \\ &= [f^3 + 3f^2(1-f)] p_{\mathtt{a}} + [f^3 + 3f^2(1-f)] p_{\mathtt{b}} \\ &= f^3 + 3f^2(1-f) = 0.028 \end{aligned}$$

So the *error* has dropped from 0.1 to 0.028
but so has the *rate*: from 1 symbol/bit to 1/3 symbol/bit.

A Simple Coding Scheme

Suppose $\mathbf{s} \in \{\mathtt{a}, \mathtt{b}\}$ and we encode by $\mathtt{a} \to \mathtt{000}$ and $\mathtt{b} \to \mathtt{111}$.
To decode we count the number of 1s and 0s and set all bits to the majority count to determine $\mathbf{s}$

$$\underbrace{\mathtt{000}, \mathtt{001}, \mathtt{010}, \mathtt{100}}_{A} \to \mathtt{a} \quad \text{and} \quad \underbrace{\mathtt{111}, \mathtt{110}, \mathtt{101}, \mathtt{011}}_{B} \to \mathtt{b}$$

If the channel Q is binary symmetric with $f = 0.1$ again

$$\begin{aligned} P(\mathbf{s}_{in} \neq \mathbf{s}_{out}) &= P(\mathbf{y} \in B | 000)\, p_{\mathrm{a}} + P(\mathbf{y} \in A | 111)\, p_{\mathrm{b}} \\ &= [f^3 + 3f^2(1-f)] p_{\mathrm{a}} + [f^3 + 3f^2(1-f)] p_{\mathrm{b}} \\ &= f^3 + 3f^2(1-f) = 0.028 \end{aligned}$$

So the *error* has dropped from 0.1 to 0.028
but so has the *rate*: from 1 symbol/bit to 1/3 symbol/bit.
Can we make the error arbitrarily small without the rate going to zero?

Channel Capacity

A key quantity when using a channel is the mutual information between its inputs X and outputs Y:

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

This measures how much what was received *reduces uncertainty* about what was transmitted.

Channel Capacity

A key quantity when using a channel is the mutual information between its inputs X and outputs Y:

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

This measures how much what was received *reduces uncertainty* about what was transmitted.

Examples (See MacKay §9.5)

- For noiseless channel $H(X|Y) = 0$ so $I(X;Y) = H(X)$.
  If $\mathbf{p}_X = (0.9, 0.1)$ then $I(X;Y) = 0.47$ bits.

Channel Capacity

A key quantity when using a channel is the mutual information between its inputs X and outputs Y:

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

This measures how much what was received *reduces uncertainty* about what was transmitted.

Examples (See MacKay §9.5)

- For noiseless channel $H(X|Y) = 0$ so $I(X;Y) = H(X)$.
  If $\mathbf{p}_X = (0.9, 0.1)$ then $I(X;Y) = 0.47$ bits.
- For binary symmetric channel with $f = 0.15$ and $\mathbf{p}_X$ as above we have $H(Y) = 0.76$ and $H(Y|X) = 0.61$ so $I(X;Y) = 0.15$ bits

Channel Capacity

A key quantity when using a channel is the mutual information between its inputs X and outputs Y:

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

This measures how much what was received *reduces uncertainty* about what was transmitted.

Examples (See MacKay §9.5)

- For noiseless channel $H(X|Y) = 0$ so $I(X;Y) = H(X)$.
  If $\mathbf{p}_X = (0.9, 0.1)$ then $I(X;Y) = 0.47$ bits.
- For binary symmetric channel with $f = 0.15$ and $\mathbf{p}_X$ as above we have $H(Y) = 0.76$ and $H(Y|X) = 0.61$ so $I(X;Y) = 0.15$ bits
- For Z channel with $f = 0.15$ and same $\mathbf{p}_X$ we have $H(Y) = 0.42$, $H(Y|X) = 0.061$ so $I(X;Y) = 0.36$ bits

Channel Capacity

A key quantity when using a channel is the mutual information between its inputs X and outputs Y:

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

This measures how much what was received *reduces uncertainty* about what was transmitted.

Examples (See MacKay §9.5)

- For noiseless channel $H(X|Y) = 0$ so $I(X;Y) = H(X)$.
  If $\mathbf{p}_X = (0.9, 0.1)$ then $I(X;Y) = 0.47$ bits.
- For binary symmetric channel with $f = 0.15$ and $\mathbf{p}_X$ as above we have $H(Y) = 0.76$ and $H(Y|X) = 0.61$ so $I(X;Y) = 0.15$ bits
- For Z channel with $f = 0.15$ and same $\mathbf{p}_X$ we have $H(Y) = 0.42$, $H(Y|X) = 0.061$ so $I(X;Y) = 0.36$ bits

Channel Capacity

A key quantity when using a channel is the mutual information between its inputs X and outputs Y:

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

This measures how much what was received *reduces uncertainty* about what was transmitted.

Examples (See MacKay §9.5)

- For noiseless channel $H(X|Y) = 0$ so $I(X;Y) = H(X)$.
  If $\mathbf{p}_X = (0.9, 0.1)$ then $I(X;Y) = 0.47$ bits.
- For binary symmetric channel with $f = 0.15$ and $\mathbf{p}_X$ as above we have $H(Y) = 0.76$ and $H(Y|X) = 0.61$ so $I(X;Y) = 0.15$ bits
- For Z channel with $f = 0.15$ and same $\mathbf{p}_X$ we have $H(Y) = 0.42$, $H(Y|X) = 0.061$ so $I(X;Y) = 0.36$ bits

So, intuitively, the reliability is "noiseless > Z > symmetric"

Channel Capacity

The mutual information measure for a channel depends on the choice of input distribution $\mathbf{p}_X$. If $H(X)$ is small then $I(X;Y) \leq H(X)$ is small.

Channel Capacity

The mutual information measure for a channel depends on the choice of input distribution $\mathbf{p}_X$. If $H(X)$ is small then $I(X;Y) \leq H(X)$ is small.

The *largest possible* reduction in uncertainty achievable across a channel is its **capacity**.

Channel Capacity

The capacity C of a channel Q is the largest mutual information between its input and output for any choice of input ensemble. That is,

$$C = \max_{\mathbf{p}_X} I(X;Y)$$

Example: For binary symmetric channel ($f = .15$), $I(X;Y)$ is maximal for $\mathbf{p}_X = (0.5, 0.5)$, so $C = 0.39$ bits (cf. $I(X;Y) = 0.15$ for $\mathbf{p}_X = (0.9, 0.1)$)

Channel Capacity

The mutual information measure for a channel depends on the choice of input distribution $\mathbf{p}_X$. If $H(X)$ is small then $I(X;Y) \leq H(X)$ is small.

The *largest possible* reduction in uncertainty achievable across a channel is its **capacity**.

Channel Capacity

The capacity C of a channel Q is the largest mutual information between its input and output for any choice of input ensemble. That is,

$$C = \max_{\mathbf{p}_X} I(X;Y)$$

Example: For binary symmetric channel ($f = .15$), $I(X;Y)$ is maximal for $\mathbf{p}_X = (0.5, 0.5)$, so $C = 0.39$ bits (cf. $I(X;Y) = 0.15$ for $\mathbf{p}_X = (0.9, 0.1)$)

Later, we will see that the capacity determines the rate at which we can communicate across a channel with arbitrarily small error.

Summary and Reading

Main Points:

- Modelling Noisy Channels
  - Noiseless, Overlap, Symmetric, Z, Erasure
- Simple Coding via Repetition
  - Probability of Error vs Transmission Rate
- Channel Capacity

Reading:

- MacKay §9.1 - §9.5
- Cover & Thomas §7.1 - §7.3

Summary and Reading

Main Points:

- Modelling Noisy Channels
  - Noiseless, Overlap, Symmetric, Z, Erasure
- Simple Coding via Repetition
  - Probability of Error vs Transmission Rate
- Channel Capacity

Reading:

- MacKay §9.1 - §9.5
- Cover & Thomas §7.1 - §7.3

Next Time: Block Coding and the Source Coding Theorem