COMP2610/6261 - Information Theory

Lecture 18: Channel Capacity

RObert C. Williamson

Research School of Computer Science

Australian National University

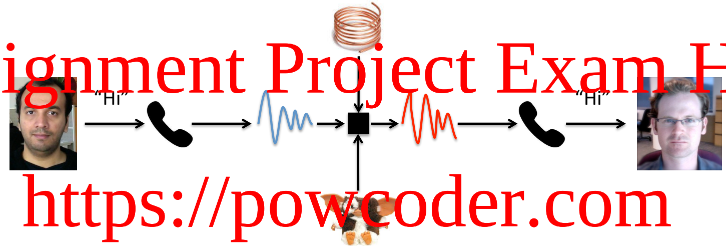
8 October, 2018

1. Channel Capacity

2. Computing Capacities

3. Summary

Channels: Recap



Source : Aditya

Encoder : Telephone handset

Channel : Analogue telephone line

Decoder : Telephone handset

Destination : Mark

Channels: Recap

A **discrete channel** Q consists of:

- an *input alphabet* $\mathcal{X} = \{a_1, \ldots, a_I\}$
- an *output alphabet* $\mathcal{Y} = \{b_1, \ldots, b_J\}$
- *transition probabilities* $P(y|x)$.

The channel Q can be expressed as a matrix

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

This represents the probability of observing b_j given that we transmit a_i

The Binary Noiseless Channel

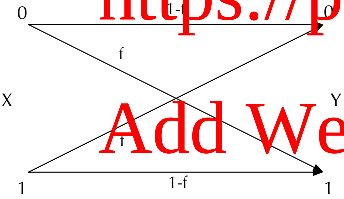
One of the simplest channels is the **Binary Noiseless Channel**. The received symbol is always equal to the transmitted symbol – there is no probability of error, hence *noiseless*.

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

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

The Binary Symmetric Channel

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

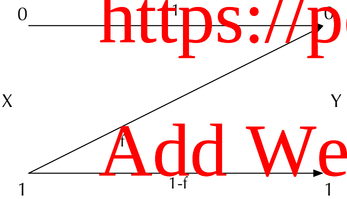


Inputs $\mathcal{X} = \{0, 1\}$; Outputs $\mathcal{Y} = \{0, 1\}$;
Transition probabilities with $P(\text{flip}) = f$

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

The Z Channel

Symbols may be corrupted over the channel asymmetrically.



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

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

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.

$$\boxed{\mathbf{s}_{in}} \xrightarrow{\text{Encoder}} \boxed{\mathbf{x}} \xrightarrow{Q} \boxed{\mathbf{y}} \xrightarrow{\text{Decoder}} \boxed{\mathbf{s}_{out}}$$

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

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

1. Channel Capacity

2. Computing Capacities

3. Summary

Mutual Information for a Channel

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

This requires we specify some particular $p(X)$

- A channel is only specified by its transition matrix!

Mutual Information for a Channel: Example

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.

Mutual Information for a Channel: Example

For binary symmetric channel with $f = 0.15$ and $\mathbf{p}_X = (0.9, 0.1)$ we have

$$
\begin{aligned}
p(Y = 1) &= p(Y = 1 \mid X = 1) \cdot p(X = 1) + p(Y = 1 \mid X = 0) \cdot p(X = 0) \\
&= (1 - f) \cdot 0.1 + f \cdot 0.9 \\
&= 0.085 + 0.135 \\
&= 0.22,
\end{aligned}
$$

and so $H(Y) = 0.76$

Further, $H(Y \mid X = 0) = H(Y \mid X = 1) = H(0.15) = 0.61$.

So, $I(X; Y) = 0.15$ bits

Mutual Information for a Channel: Example

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.

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,

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

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

Computing Capacities

Definition of capacity for a channel Q with inputs $\mathcal{A}_X$ and ouputs $\mathcal{A}_Y$:

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

How do we actually calculate this quantity?

1. Compute the mutual information $I(X;Y)$ for a general $\mathbf{p}_X$
2. Determine which choice of $\mathbf{p}_X$ maximises $I(X;Y)$
3. Use that maximising value to determine C

Binary Symmetric Channel:

We first consider the *binary symmetric channel* with $\mathcal{A}_X = \mathcal{A}_Y = \{0,1\}$ and flip probability f. It has transition matrix

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

Computing Capacities

Binary Symmetric Channel - Step 1

The mutual information can be expressed as $I(X;Y) = H(Y) - H(Y|X)$. We therefore need to compute two terms: $H(Y)$ and $H(Y|X)$ so we need the distributions $P(y)$ and $P(y|x)$.

Computing $H(Y)$:

- $P(y=0) = (1-f) \cdot P(x=0) + f \cdot P(x=1) = (1-f) \cdot p_0 + f \cdot p_1$
- $P(y=1) = (1-f) \cdot P(x=1) + f \cdot P(x=0) = f \cdot p_0 + (1-f) \cdot p_1$

In general, $\mathbf{q} := \mathbf{p}_Y = Q\mathbf{p}_X$, so above calculation is just

$$\mathbf{q} = \mathbf{p}_Y = \begin{bmatrix} 1-f & f \\ f & 1-f \end{bmatrix} \begin{bmatrix} p_0 \\ p_1 \end{bmatrix}$$

Using $H_2(q) = -q \log_2 q - (1-q) \log_2(1-q)$ and letting $q = q_1 = P(y=1)$ we see the entropy

$$H(Y) = H_2(q_1) = H_2(f \cdot p_0 + (1-f) \cdot p_1)$$

Computing Capacities

Binary Symmetric Channel - Step 1

Computing $H(Y|X)$:

Since $P(y|x)$ is described by the matrix Q, we have

$H(Y|x=0) = H_2(P(y=1|x=0)) = H_2(Q_{1,0}) = H_2(f)$

and similarly,

$H(Y|x=1) = H_2(P(y=1|x=1)) = H_2(Q_{0,1}) = H_2(f)$

So,

$$H(Y|X) = \sum_x H(Y|x)P(x) = \sum_x H_2(f)P(x) = H_2(f)\sum_x P(x) = H_2(f)$$

Computing $I(X;Y)$:

Putting it all together gives

$$\boxed{I(X;Y) = H(Y) - H(Y|X) = H_2(f\cdot p_0 + (1-f)\cdot p_1) - H_2(f)}$$

Computing Capacities

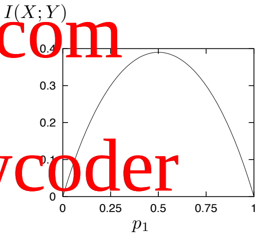
Binary Symmetric Channel - Steps 2 and 3

Binary Symmetric Channel (BSC) with flip probability $f \in [0, 1]$:

$$I(X; Y) = H_2(fp_0 + (1-f)p_1) - H_2(f)$$

Examples:

- BSC ($f = 0$) and $\mathbf{p}_X = (0.5, 0.5)$:
  $I(X; Y) = H_2(0.5) - H_2(0) = 1$
- BSC ($f = 0.15$) and $\mathbf{p}_X = (0.5, 0.5)$:
  $I(X; Y) = H_2(0.5) - H_2(0.15) \approx 0.39$
- BSC ($f = 0.15$) and $\mathbf{p}_X = (0.9, 0.1)$:
  $I(X; Y) = H_2(0.22) - H_2(0.15) \approx 0.15$



$I(X; Y), f = 0.15$

Maximise $I(X; Y)$: Since $I(X; Y)$ is symmetric in p_1 it is maximised when $p_0 = p_1 = 0.5$ in which case $C = 0.39$ for BSC with $f = 0.15$.

For a binary symmetric channel, we could also argue

$$
\begin{aligned}
I(X;Y) &= H(Y) - H(Y \mid X) \\
&= H(Y) - \sum_x p(X = x) \cdot H(Y \mid X = x) \\
&= H(Y) - \sum_x p(X = x) \cdot H(f) \\
&= H(Y) - H(f) \\
&\leq 1 - H(f),
\end{aligned}
$$

where equality of the last line holds for **uniform** $\mathbf{p}_X$

Symmetric Channels

The BSC was easy to work with due to considerable symmetry

Symmetric Channel

A channel with input $\mathcal{A}_X$ and outputs $\mathcal{A}_Y$ and matrix Q is **symmetric** if $\mathcal{A}_Y$ can be partitioned into subsets $Y' \subseteq Y$ so that each sub-matrix Q' containing only rows for outputs Y' has:

- Columns that are all permutations of each other
- Rows that are all permutations of each other

Symmetric Channels: Examples

$\mathcal{A}_X = \mathcal{A}_Y = \{0, 1\}$

$$Q = \begin{bmatrix} 0.9 & 0.1 \\ 0.1 & 0.9 \end{bmatrix}$$

Symmetric

Subsets: $\{0, 1\}$

$\mathcal{A}_X = \{0, 1\}, \mathcal{A}_Y = \{0, ?, 1\}$

$$Q = \begin{bmatrix} 0.7 & 0.1 \\ 0.2 & 0.2 \\ 0.1 & 0.7 \end{bmatrix}$$

Symmetric

Subsets: $\{0, 1\}$, $\{?\}$

$\mathcal{A}_X = \mathcal{A}_Y = \{0, 1\}$

$$Q = \begin{bmatrix} 0.9 & 0 \\ 0.1 & 1 \end{bmatrix}$$

Not Symmetric

If one of our partitions has just one row, then every element in that must be equal for the columns to be permutations of each other

Simplest case: all rows and columns are permutations of each other

- But this is not a requirement

Channel Capacity for Symmetric Channels

For symmetric channels, the optimal distribution for the capacity has a simple form:

Theorem

If Q is symmetric, then its capacity is achieved by a uniform distribution over $\mathcal{X}$.

Exercise 10.10 in MacKay

Computing Capacities in General

What can we do if the channel is not symmetric?

- We can still calculate $I(X;Y)$ for a general input distribution $\mathbf{p}_X$
- Finding the maximising $\mathbf{p}_X$ is more challenging

What to do once we know $I(X;Y)$?

- $I(X;Y)$ is *concave* in $\mathbf{p}_X \implies$ single maximum
- For *binary* inputs, just look for stationary points (not for $|\mathcal{A}_X| > 2$) i.e., where $\frac{d}{dp} I(X;Y) = 0$ for $\mathbf{p}_X = (1-p, p)$
- In general, need to consider distributions that place 0 probability on one of the inputs

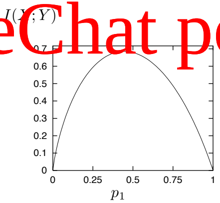
Computing Capacities in General

Example (Z Channel with $P(y=0|x=1)=f$):

$$H(Y) = H_2(P(y=1)) = H_2(0p_0 + (1-f)p_1)$$
$$= H_2((1-f)p_1)$$

$$H(Y|X) = p_0 H_2(P(y=1|x=0)) + p_1 H_2(P(y=0|x=1))$$
$$= p_0 \underbrace{H_2(0)}_{=0} + p_1 H_2(f)$$

$$I(X;Y) = H_2((1-f)p_1) - p_1 H_2(f)$$



$I(X;Y)$ for Z channel with $f=0.15$

Computing Capacities in General

Example (Z Channel):
Showed earlier that $I(X;Y) = H_2((1-f)p) - pH_2(f)$ so solve

$$\frac{d}{dp} I(X;Y) = 0 \iff (1-f)\log_2\left(\frac{1-(1-f)p}{(1-f)p}\right) - H_2(f) = 0$$

$$\iff \frac{1-(1-f)p}{(1-f)p} = 2^{H_2(f)/(1-f)}$$

$$\iff p = \frac{1/(1-f)}{1+2^{H_2(f)/(1-f)}}$$

For $f = 0.15$, we get $p = \frac{1/0.85}{1+2^{0.61/0.85}} \approx 0.44$ and so
$C = H_2(0.38) - 0.44 H_2(0.15) \approx 0.685$

Homework: Show that $\frac{d}{dp}H_2(p) = \log_2 \frac{1-p}{p}$

Why Do We Care?

We have a template for computing channel capacity for generic channels

But what does this tell us?

- How, if at all, does it relate to the error probability when decoding?
- What, if anything, does it tell us about the amount of redundancy we can get away with when encoding?

We will see next time that there is a deep connection between the capacity and the best achievable rate of transmission

- Rates above the capacity cannot be achieved while ensuring arbitrarily small error probabilities

Summary and Conclusions

Mutual information between input and output should be large

- Depends on input distribution

Capacity is the maximal possible mutual information

Can compute easily for symmetric channels

- Can compute explicitly for generic channels