Information theory

Discrete channel capacity

Georgios Ropokis

CentaraleSupélec, Campus Rennes

Table of contents

Definition of a discrete memoryless channel and its information capacity

Defining communication

- Definition of *successful* communication: We say that communication between transmitter A and receiver B is successful if both agree on what was the message that was sent.
- Definition of a (discrete) communication channel: The combination of the following components:
    - A discrete input alphabet $\mathcal{X}$
    - A discrete output alphabet $\mathcal{Y}$
    - A probability transition matrix $p\left(y|x\right)$ that gives the probability of observing the output symbol $y \in \mathcal{Y}$, given that symbol $x \in \mathcal{Y}$ was sent
- Memoryless channel: The output distribution does not depend on previous channel inputs or outputs. Ony the input of the current time instance influences the distribution of the current output.

Information channel capacity

- Definition: We define the information channel capacity of a discrete memoryless channel as the the quantity:

$$C = \max_{p(x)} I\left(X; Y\right)$$

  where maximum is taken over all possible input probability distributions $p\left(x\right)$.
- Note: Finding the information capacity of the channel corresponds to finding the probability distribution of the input symbols that maximizes the mutual information.

The Binary symmetric channel

- For each time slot, a bit is transmitted, where p_0 is the probability that the bit to be transmitted has a value equal to 0 and $p_1 = 1 - p_0$ is the probability that the bit to be transmitted is equal to 1.
- During the transmission, errors occur in a symmetric manner, i.e., $\Pr(Y = 1|X = 0) = \Pr(Y = 0|X = 1) = q$.
- Entropy of Y:

$$\begin{aligned} H(Y) &= -\Pr(Y=0)\log(\Pr(Y=0)) - \Pr(Y=1)\log(\Pr(Y=1)) \\ &= -\left(p_0(1-q)+p_1 q\right)\log\left(\left(p_0(1-q)+p_1 q\right)\right) \\ &\quad -\left(p_1(1-q)+p_0 q\right)\log\left(p_1(1-q)+p_0 q\right) \end{aligned}$$

- We then calculate the conditional entropy $H(Y|X)$

$$\begin{aligned} H(Y|X) &= p_0 H(Y|X=0) + (1-p_0) H(Y|X=1) \\ &= -p_0\left(q\log q + (1-q)\log(1-q)\right) \\ &\quad -(1-p_0)\left(q\log q + (1-q)\log(1-q)\right) \\ &= -\left(q\log q + (1-q)\log(1-q)\right) \end{aligned}$$

The conditional entropy is independent of p_0, p_1!

Capacity of a binary symmetric channel

We calculate the mutual information:

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

Since the capacity is the maximum (with respect to p_0, p_1) of $I(X;Y)$ and $H(Y|X)$ is independent of p_0, p_1, maximizing $I(X;Y)$ is equivalent to maximizing $H(Y)$. By introducing $\tilde{p} = p_1(1-q) + p_0 q$ this is equivalently written as:

$$H(Y) = -\tilde{p}\log\tilde{p} - (1-\tilde{p})\log(1-\tilde{p})$$

which is maximized if $\tilde{p} = \frac{1}{2}$, or equivalently if $p_0 = p_1 = \frac{1}{2}$. The capacity then becomes:

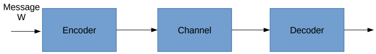
$$C = \max_{p_0} I(X;Y) = 1 + (q\log q + (1-q)\log(1-q))$$

Basic properties of channel capacity

1. $C \geq 0$, since $I\left(X;Y\right) \geq 0$.
2. $C \leq \log|\mathcal{X}|$, since $C = \max I\left(X;Y\right) \leq \max H\left(X\right) = \log|\mathcal{X}|$.
3. $C \leq \log|\mathcal{Y}|$
4. $I\left(X;Y\right)$ is a continuous function of $p\left(x\right)$.
5. $I\left(X;Y\right)$ is a concave function of $p\left(x\right)$, defined over a convex set. As a result, any local optimum is a global optimum.

Representing a communications system

Structure of a communications code



- A source producing messages $W \in \{1, \ldots, M\}$ coming from a set of M different messages.
- Encoder: produces a codeword/signal $x^n(W)$ having a length of n bits.
- Codebook: The set of all codewords corresponding to the M messages $\{x^n(1), \ldots, x^n(M)\}$
- Discrete memoryless channel extension: $(\mathcal{X}^n, p(y^n|x^n), \mathcal{Y}^n)$ where $p\left(y_k|x_k, y^{k-1}\right) = p\left(y_k|x_k\right), k = 1, \ldots, n$.
- A deterministic decoding rule $g : \mathcal{Y}^n \mapsto \{1, \ldots, n\}$.

Structure of a communications system

- The rate of the code:

$$R = \log_2 M/n \tag{1}$$

- The conditional error probability of a code:

$$\lambda_i = \Pr\left(g\left(Y^n\right) \neq i | X^n = x^n(i)\right) \tag{2}$$

- The maximal error probability $\lambda^{(n)}$:

$$\lambda^{(n)} = \max\left\{\lambda_1, \ldots, \lambda_M\right\} \tag{3}$$

Defining the capacity of a cummunications system

Shannon's channel coding theorem

For a discrete memoryless channel, all rates below the information capacity C are achievable, i.e., ther exists a sequence of $(2^{nR}, n)$ codes with maximum probability of error $\lambda^{(n)} \to 0$.