Outline Chapter 4: Information Channels

Discrete Channel Model

Channel Capacity

Binary Channels
  Binary Symmetric Channel (BSC)
  Binary Asymmetric Channel (BAC)
  Binary Z-Channel (BZC)
  Binary Asymmetric Erasure Channel (BAEC)

Channel Coding

Decoding Rules

Error Probabilities

The Noisy Coding Theorem

Converse of the Noisy Coding Theorem
Communication Channel
from an information theoretic point of view
Discrete Channel Model

Discrete information channels are described by

- A pair of random variables

\[(X, Y)\] with support \(\mathcal{X} \times \mathcal{Y}\),

- \(X\) is the input r.v., \(\mathcal{X} = \{x_1, \ldots, x_m\}\) the input alphabet.
- \(Y\) is the output r.v., \(\mathcal{Y} = \{y_1, \ldots, y_d\}\) the output alphabet.

- The channel matrix

\[W = (w_{ij})_{i=1,\ldots,m, j=1,\ldots,d}\]

with

\[w_{ij} = P(Y = y_j \mid X = x_i), \quad i = 1, \ldots, m, \quad j = 1, \ldots, d\]

- Input distribution

\[P(X = x_i) = p_i, \quad i = 1, \ldots, m,\]

\[p = (p_1, \ldots, p_m).\]
Lemma 4.1

\[ H(Y) = H(pW) \]
\[ H(Y | X = x_i) = H(w_i) \]
\[ H(Y | X) = \sum_{i=1}^{m} p_i H(w_i) \]
Channel Capacity

Mutual information

\[ I(X; Y) = H(Y) - H(Y \mid X) \]
\[ = H(pW) - \sum_{i=1}^{m} p_i H(w_i) \]
\[ = \sum_{i=1}^{m} p_i D(w_i \parallel pW) = I(p; W), \]

\( D \) denoting the Kulback-Leibler divergence.

**Aim:** For a given channel \( W \) use the input distribution that maximizes mutual information \( I(X; Y) \).

**Definition 4.2.**

\[ C = \max_{(p_1, \ldots, p_m)} I(X; Y) = \max_{p} I(p; W) \]

is called *channel capacity*.

Determining capacity is in general a complicated optimization problem.
Binary Symmetric Channel (BSC)

Example 4.3.

Input distribution \( p = (p_0, p_1) \)

Channel matrix

\[
W = \begin{pmatrix}
1 - \varepsilon & \varepsilon \\
\varepsilon & 1 - \varepsilon
\end{pmatrix}
\]

Mutual Information:

\[
I(X; Y) = I(p; W) = H(pW) - \sum_{i=1}^{m} p_i H(w_i)
\]

\[
= H(p_0(1 - \varepsilon) + p_1 \varepsilon, \varepsilon p_0 + (1 - \varepsilon) p_1) - H(\varepsilon, 1 - \varepsilon)
\]

The maximum of \( I(p, W) \) over all input distributions \( (p_0, p_1) \) is achieved at

\( (p_0^*, p_1^*) = (0.5, 0.5) \)
Binary Symmetric Channel (BSC)

Hence, $p_0^* = p_1^* = \frac{1}{2}$ is capacity-achieving. It holds

$$C = \max_{(p_0, p_1)} I(X; Y) = 1 + (1 - \epsilon) \log_2(1 - \epsilon) + \epsilon \log_2 \epsilon$$

Capacity of the BSC as a function of $\epsilon$:
Channel Capacity (ctd.)

Given a channel with channel matrix $W$. To compute channel capacity solve

$$C = \max_p l(p; W) = \max_p \sum_{i=1}^{m} p_i D(w_i \| pW)$$

**Theorem 4.4.**
The capacity of the channel $W$ is attained at $p^* = (p_1^*, \ldots, p_m^*)$ if and only if

$$D(w_i \| p^*W) = \zeta \text{ for all } i = 1, \ldots, m.$$ 

for all $i = 1, \ldots, m$ with $p_i > 0$.

Moreover,

$$C = l(p^*; W) = \zeta.$$
Channel Capacity (ctd.)

Proof of the Theorem:

Mutual information \( I(p; W) \) is a concave function of \( p \). Hence the KKT conditions (cf., e.g., Boyd and Vandenberge 2004) are necessary and sufficient for optimality of some input distribution \( p \). Using the above representation some elementary algebra shows that

\[
\frac{\partial}{\partial p_i} I(p; W) = D(w_i \| pW) - 1.
\]

The full set of KKT conditions now reads as

\[
\sum_{j=1}^{m} p_j = 1 \\
p_i \geq 0, \ i = 1, \ldots, m \\
\lambda_i \geq 0, \ i = 1, \ldots, m \\
\lambda_i p_i = 0, \ i = 1, \ldots, m \\
D(w_i \| pW) + \lambda_i + \nu = 0, \ i = 1, \ldots, m
\]

which shows the assertion.
Channel Capacity (ctd.)

Denote self information by \( \rho(q) = -q \log q, \quad q \geq 0. \)

**Theorem 4.5.** (G. Alirezaei, 2018)

Given a channel with square channel matrix \( \mathbf{W} = (w_{ij})_{i,j=1,...,m}. \) Assume that \( \mathbf{W} \) is invertible with inverse

\[
T = (t_{ij})_{i,j=1,...,m}.
\]

Then, measured in nats, the capacity is

\[
C = \ln \left( \sum_{k} e^{\exp\left\{- \sum_{i,j} t_{ki} \rho(w_{ij}) \right\}} \right)
\]

and the capacity achieving distribution is given by

\[
p_{\ell}^* = e^{-C} \sum_{k} t_{ks} e^{\exp\left\{- \sum_{i,j} t_{ki} \rho(w_{ij}) \right\}} = \frac{\sum_{k} t_{ks} e^{\exp\left\{- \sum_{i,j} t_{ki} \rho(w_{ij}) \right\}}}{\sum_{k} e^{\exp\left\{- \sum_{i,j} t_{ki} \rho(w_{ij}) \right\}}}
\]
Binary Asymmetric Channel (BAC)

Example 4.6.

The capacity-achieving distribution is

\[ p_0^* = \frac{1}{1+b}, \quad p_1^* = \frac{b}{1+b}, \]

with

\[ b = \frac{a\varepsilon - (1-\varepsilon)}{\delta - a(1-\delta)} \quad \text{and} \quad a = \exp \left( \frac{h(\delta) - h(\varepsilon)}{1-\varepsilon-\delta} \right), \]

and \( h(\varepsilon) = H(\varepsilon, 1-\varepsilon) \), the entropy of \((\varepsilon, 1-\varepsilon)\).

Note that \( \varepsilon = \delta \) yields the previous result for the BSC.
Binary Asymmetric Channel (BAC)

Derivation of capacity for the BAC:

By Theorem 4.4 the capacity-achieving input distribution $p = (p_0, p_1)$ satisfies

$$D(w_1||pW) = D(w_2||pW).$$

This is an equation in the variables $p_0, p_1$ which jointly with the condition $p_0 + p_1 = 1$ has the solution

$$p_0^* = \frac{1}{1 + b}, \quad p_1^* = \frac{b}{1 + b},$$

with

$$b = \frac{a\epsilon - (1 - \epsilon)}{\delta - a(1 - \delta)} \quad \text{and} \quad a = \exp \left( \frac{h(\delta) - h(\epsilon)}{1 - \epsilon - \delta} \right),$$

and $h(\epsilon) = H(\epsilon, 1 - \epsilon)$, the entropy of $(\epsilon, 1 - \epsilon)$. 

Rudolf Mathar, Information Theory, RWTH Aachen, WS 2018/19
Example 4.7.
The so called Z-channel is a special case of the BAC with $\varepsilon = 0$.

The capacity-achieving distribution is obtained from the BAC by setting $\varepsilon = 0$. 

The capacity-achieving distribution is obtained from the BAC by setting $\varepsilon = 0$. 
Example 4.8.

\[
W = \begin{pmatrix}
1 - \varepsilon & \varepsilon & 0 \\
0 & \delta & 1 - \delta
\end{pmatrix}
\]

The capacity-achieving distribution is determined by finding the solution \( x^* \) of

\[
\varepsilon \log \varepsilon - \delta \log \delta = (1 - \delta) \log(\delta + \varepsilon x) - (1 - \varepsilon) \log(\varepsilon + \delta/x)
\]

and setting

\[
\frac{p_0^*}{p_1^*} = x^*, \quad p_0^* + p_1^* = 1.
\]
Binary Asymmetric Erasure Channel (BAEC)

Derivation of capacity for the BAEC:

By Theorem 4.4 the capacity-achieving distribution $p^* = (p_0^*, p_1^*)$, $p_0^* + p_1^* = 1$ is given by the solution of

\[
(1 - \varepsilon) \log \frac{1 - \varepsilon}{p_0(1 - \varepsilon)} + \varepsilon \log \frac{\varepsilon}{p_0\varepsilon + p_1\delta}
= \delta \log \frac{\delta}{p_0\varepsilon + p_1\delta} + (1 - \delta) \log \frac{1 - \delta}{p_0(1 - \delta)},
\]

(2)

Substituting $x = \frac{p_0}{p_1}$, equation (2) reads equivalently as

\[
\varepsilon \log \varepsilon - \delta \log \delta = (1 - \delta) \log(\delta + \varepsilon x) - (1 - \varepsilon) \log(\varepsilon + \delta/x)
\]

By differentiating w.r.t. $x$ it is easy to see that the right hand side is monotonically increasing such that exactly one solution $p^* = (p_1^*, p_2^*)$ exists, which can be numerically computed.