### I.7 Classical channels and channel codes

**Communication system:**

\[ m \xrightarrow{\text{encoder}} x^n \xrightarrow{\text{noisy channel}} y^n \xrightarrow{\text{decoder}} \hat{m} \]

In the data compression scenario, there was no noise, i.e., \( y^n = x^n \), and the goal of the encoder/compressor was to remove redundancy from the message to compress it as much as possible. When there is a noisy transmission channel, then the encoder should add sufficient redundancy such that decoding after noise is possible (with small error probability).

We model the noise as follows:

**Definition:** A discrete channel between finite sets \( X \) and \( Y \) is described by a linear map \( T : \mathbb{R}^{|X|} \to \mathbb{R}^{|Y|} \) that maps probability distributions on \( X \) to probability distributions on \( Y \); i.e., the characterising matrix \( T \in \mathbb{R}^{|Y| \times |X|} \) is a stochastic matrix, i.e., satisfies \( \sum_{y \in Y} T_{yx} = 1 \) \( \forall x \in X \).

For any \( x \in X \) and \( y \in Y \), \( T_{yx} =: p(y|x) \) is interpreted as the probability of \( y \) appearing at the channel output if the input was \( x \). Thus, \( p(y|x) \) are conditional probabilities.

\[ \text{input} \quad x \xrightarrow{T} y \in Y \quad \text{output} \]
The "discrete memoryless extension" to a map of the discrete channel $T$ is the discrete channel $T^\otimes n : R_+^n \rightarrow R_+^n$, where $(T^\otimes n)_{yx} = \prod_{i=1}^n T_{y_i|x_i} \quad \text{for } x = (x_1, \ldots, x_n) \in \chi^n \text{ and } y = (y_1, \ldots, y_n) \in \gamma^n$.

Thus, the transition probabilities do not depend on previous inputs or outputs.

Unless stated otherwise, by "channel" we will mean a discrete channel and its discrete memoryless extension to $n$ uses.

**Examples:**

- **Binary symmetric channel**: $T = \begin{pmatrix} 1-p & p \\ p & 1-p \end{pmatrix}$, $p \in [0,1]$

  - Input distribution when sending 0 is $\vec{\pi} = (\frac{1}{2})$
  - Output distribution $\vec{q} = T\vec{\pi} = \begin{pmatrix} 1-p & p \\ p & 1-p \end{pmatrix} \begin{pmatrix} \frac{1}{2} \\ \frac{1}{2} \end{pmatrix} = \begin{pmatrix} \frac{1}{2}(1-p) \\ \frac{1}{2}(1-p) \end{pmatrix}$

  - Input distribution $\vec{q} = (\frac{1}{4})$ leads to output distribution $\vec{q} = T\vec{q} = \begin{pmatrix} 1-p & p \\ p & 1-p \end{pmatrix} \begin{pmatrix} \frac{1}{4} \\ \frac{1}{4} \end{pmatrix} = \begin{pmatrix} \frac{1}{4}(1-p) + \frac{1}{4}p \\ \frac{1}{4}(1-p) + \frac{1}{4}p \end{pmatrix}$

- **Binary erasure channel**: $T = \begin{pmatrix} 1-p & 0 \\ p & 1-p \end{pmatrix}$, $p \in [0,1]$

  - Input 0 is lost with probability $p$.

**Remarks:** The linearity requirement on $T : R_+^1 \rightarrow R_+^1$ can be justified as follows:

For any probability distribution $\pi$ on $X$ and any $x \in \chi$, the encoder's uncertainty about which distribution the input follows, it should hold that $T(\lambda \pi + (1-\lambda)x) = \lambda T(\pi) + (1-\lambda) T(x)$, since the channel should not be affected by the ignorance $\lambda$. Then, when embedding the set of probability distributions into a linear space $R_+^1$, the map $T$ has a linear extension.
We want to send information (messages) using the channel many times:

**Definition:** A "(M,n) code" (with M, n ∈ M) for a channel \( T : \mathbb{R}_+^{1 \times 1} \rightarrow \mathbb{R}_+^{1 \times 1} \)
consists of:

(i) an index set \( M \) ("set of messages") with \( |M| = M \),
(ii) an encoding function \( x : M \rightarrow \mathbb{X}^n \),
(iii) a decoding function \( g : \mathbb{Y}^n \rightarrow M \).

\( n \) is called the "blocklength" of the code, \( x(m) \) for \( m \in M \) are the "codewords", and \( x(M) \) is the "codebook".

**Example:** repetition code (e.g., for the binary symmetric channel)

\[ M = 2, \quad M = \{0, 1\}, \quad n = 3 \]

\[ x : 0 \rightarrow 000, \quad 1 \rightarrow 111 \]

\[ g : 000, 001, 010, 011, 100, 110, 101, 111 \rightarrow 0 \quad \text{decoding by} \]

\[ 111, 110, 101, 101, 011 \rightarrow 1 \] "majority vote"

**Error:** For a \((M,n)\) code for channel \( T \):

- **error probability** for message \( m \in M \):
  \[
  \lambda_m := \text{prob} [ g(Y^n) \neq m \mid X^n = x(m) ] = \sum_{y : g(y) \neq m} p(y \mid x(m))
  \]

  where \( p(y \mid x) = \prod_{i=1}^{n} p(y_i \mid x_i) \).

  to channel \( T \).

- **maximal error probability:** \( \lambda_m := \max_{m \in M} \lambda_m \)

- **average error probability:** \( \lambda_2 := \frac{1}{M} \sum_{m \in M} \lambda_m \)