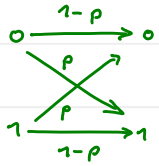
IV. Shannon's noisy channel coding theorem

IV.1. Discrete memoryless channels

Def.: Let $\mathcal{X}$ and $\mathcal{Y}$ be finite sets. A map $S : \mathbb{R}_+^{|\mathcal{X}|} \to \mathbb{R}_+^{|\mathcal{Y}|}$ describes a "discrete memoryless channel" and the characterizing matrix $S \in \mathbb{R}_+^{|\mathcal{Y}| \times |\mathcal{X}|}$ is a "stochastic matrix" if $S_{yx} =: p(y|x)$ are conditional probabilities, i.e., $\forall_{x \in \mathcal{X}} : \sum_{y \in \mathcal{Y}} p(y|x) = 1$.

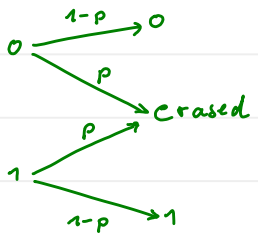
remarks:
- "memoryless" refers to the fact that n uses of the channel will be described by $S^{\otimes n} := S \otimes \dots \otimes S : \mathbb{R}_+^{|\mathcal{X}^n|} \to \mathbb{R}_+^{|\mathcal{Y}^n|}$
  where $\left(S^{\otimes n}\right)_{y,x} = \prod_{i=1}^{n} p(y_i|x_i)$, $x \in \mathcal{X}^n$, $y \in \mathcal{Y}^n$
  That is, the transition probabilities do not depend on what was sent in the past.
- in the following "channel" is meant to be discrete & memoryless

Examples:
- binary symmetric channel: $S = \begin{pmatrix} 1-p & p \\ p & 1-p \end{pmatrix}$, $p \in [0,1]$



"bit flip" occurs with prob. p

- binary erasure channel: $S = \begin{pmatrix} 1-p & 0 \\ p & p \\ 0 & 1-p \end{pmatrix}$



bit is lost with prob. p

IV.2. Codes, errors and rates

Def.: An "(M,n) code" for a channel S with input alphabet $\mathcal{X}$ and output alphabet $\mathcal{Y}$ consists of

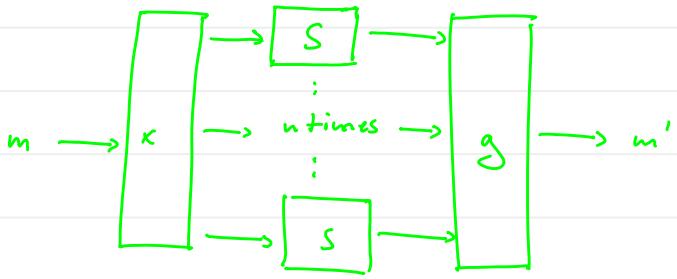
(i) an index set $\mathcal{M}$ (= set of messages) with $|\mathcal{M}| = M$

(ii) an encoding function $x : \mathcal{M} \to \mathcal{X}^n$

with "codewords" $x(m)$, $m \in \mathcal{M}$ and "codebook" $x(\mathcal{M})$

(iii) a decoding function $g : \mathcal{Y}^n \to \mathcal{M}$

n is called "blocklength"



Example: repetition code: $x : \{0,1\} \to \{0,1\}^3$, $0 \mapsto 000$, $1 \mapsto 111$

$g : \{0,1\}^3 \to \{0,1\}$ by majority vote

Errors:

- conditional prob. of error: $\lambda_m := \sum_{y : g(y) \neq m} p(y \mid x(m))$, $m \in \mathcal{M}$

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

- max. prob. of error: $\lambda^{(n)} := \max \{ \lambda_m \}_{m \in \mathcal{M}}$

- average prob. of error: $p_e^{(n)} := \frac{1}{M} \sum_{m \in \mathcal{M}} \lambda_m$

Def.: • The "rate" of an (M,n) code is $R := \frac{\log M}{n}$ (bits / transmission)

• A rate R is called "achievable" for a given channel iff there exists a sequence of $(\lceil 2^{nR} \rceil, n)$ codes s.t. $\lambda^{(n)} \to 0$ as $n \to \infty$.

• The "capacity" $C(S)$ of a channel S is the supremum over all achievable rates.

remark: the rate of a repetition code $0 \mapsto 0^n$, $1 \mapsto 1^n$ is $R = \frac{1}{n}$
So if we require $\lambda^{(n)} \to 0$, then $R \to 0$ for generic channels.

IV.3. Joint AEP

Def.: Let $n \in \mathbb{N}$, $\varepsilon > 0$ and $p : \mathcal{X} \times \mathcal{Y} \to \mathbb{R}$ be the joint distribution of random variables X and Y with ranges $\mathcal{X}, \mathcal{Y}$. The set of jointly typical sequences w.r.t. to the joint distribution p is defined as

$$A_\varepsilon^{(n)} := \left\{ (x,y) \in \mathcal{X}^n \times \mathcal{Y}^n \;\middle|\; \begin{array}{l} \left| -\frac{1}{n} \log p(x) - H(X) \right| < \varepsilon \\ \wedge \left| -\frac{1}{n} \log p(y) - H(Y) \right| < \varepsilon \\ \wedge \left| -\frac{1}{n} \log p(x,y) - H(X,Y) \right| < \varepsilon \end{array} \right\}$$

where $p(x,y) := \prod_{i=1}^{n} p(x_i, y_i)$ and $p(x)$ & $p(y)$ are the marginals.

Thm.: (joint AEP) Let $A_\varepsilon^{(n)}$ be the set of jointly typical sequences w.r.t. the joint distribution of X and Y. Then

1) $p(B_\varepsilon^{(n)}) > 1-\varepsilon$ for n suff. large

2) $|B_\varepsilon^{(n)}| \leq 2^{n(H(X,Y)+\varepsilon)}$ $\forall n \in \mathbb{N}$

3) $|B_\varepsilon^{(n)}| \geq (1-\varepsilon)\, 2^{n(H(X,Y)-\varepsilon)}$ for n suff. large

4) If $\tilde{X}^n, \tilde{Y}^n$ are i.i.d. random variables with individual ranges $\mathcal{X}$ and $\mathcal{Y}$ and joint distribution $P_r(\tilde{X}=x, \tilde{Y}=y) =: \tilde{p}(x,y)$ of the form $\tilde{p}(x,y) = p(x)p(y)$ where $p(x)$ and $p(y)$ are the marginal distributions of X and Y respectively. Then

a) $(1-\varepsilon)\, 2^{-n(I(X;Y)+3\varepsilon)} \leq P_r\left((\tilde{X}^n, \tilde{Y}^n) \in B_\varepsilon^{(n)}\right)$ for n suff. large,

b) $P_r\left((\tilde{X}^n, \tilde{Y}^n) \in B_\varepsilon^{(n)}\right) \leq 2^{-n(I(X;Y)-3\varepsilon)}$

proof: 1), 2) and 3) are proven in complete analogy to the AEP for $A_\varepsilon^{(n)}$.

4) a) $$P_r\left((\tilde{X}^n, \tilde{Y}^n) \in B_\varepsilon^{(n)}\right) = \sum_{(x,y)\in B_\varepsilon^{(n)}} p(x)p(y)$$
$$\leq |B_\varepsilon^{(n)}|\, 2^{-n(H(X)-\varepsilon)}\, 2^{-n(H(Y)-\varepsilon)}$$
$$\overset{2)}{\leq} 2^{-n(I(X;Y)-3\varepsilon)}$$

b) $$P_r\left((\tilde{X}^n, \tilde{Y}^n) \in B_\varepsilon^{(n)}\right) = \sum_{(x,y)\in B_\varepsilon^{(n)}} p(x)p(y)$$
$$\geq |B_\varepsilon^{(n)}|\, 2^{-n(H(X)+\varepsilon)}\, 2^{-n(H(Y)+\varepsilon)}$$
$$\overset{3)}{\geq} (1-\varepsilon)\, 2^{-n(I(X;Y)+3\varepsilon)}$$

□