Channel Coding Theorem for Discrete Memoryless Channels

Melina Woitun

11. Juni 2015

Definition (Discrete Channels)**.**

1. We denote $(\mathcal{X}, p(y|x), \mathcal{Y})$ a discrete channel, where $\mathcal{X}$ and $\mathcal{Y}$ are finite sets and for the collection of probability mass functions $p(y|x)$ (one for each $x \in \mathcal{X}$)the following holds: $\forall x, y : p(y|x) \geq 0$ and $\forall x : \sum_{y \in \mathcal{Y}} p(y|x) = 1$.

2. We call $(\mathcal{X}^n, p(y^n|x^n), \mathcal{Y}^n)$ the n-th extension of the discrete channel, where $p(y_k|x_k) = p(y_k|x^k, y^{k-1})$. For a memoryless channel y_k does not depend on the previous inputs x^{k-1} nor the previous Outputs y^{k-1}. Therefore $p(y^n|x^n) = \prod_{i=1}^{n} p(y_i|x_i)$.

In the following *channel* is meant to be discrete and memoryless.

Definition (Channel capacity)**.**

The channel capacity is $C = \max_{p(x)} I(X;Y)$,where the maximum is taken over all possible input distributions.

Lemma (Properties of the Capacity)**.**

1. $C \geq 0$

2. $C \leq \log |\mathcal{X}|$ and $C \leq \log |\mathcal{Y}|$

3. $I(X;Y)$ is continuous

4. $I(X;Y)$ is concave over a closed convex set

Definition ((M, n) code)**.**

An (M, n) code for $(\mathcal{X}, p(y|x), \mathcal{Y})$ consists of the following:

1. an index set $\{1, 2, ..., M\}$

2. an encoding function $x^n : \{1, 2, ..., M\} \to \mathcal{X}^n$, with the resulting codewords $x^n(1), x^n(2), ..., x^n(M)$

3. a decoding function $g : \mathcal{Y}^n \to \{1, 2, ..., M\}$

Definition (Probability of Errors)**.**

1. $\lambda_i = \Pr(g(Y^n) \neq i | X^n = x^n(i)) = \sum_{\substack{y^n \\ g(y^n) \neq i}} p(y^n|x^n(i))$ for $i \in \{1, 2, ..., M\}$ is the conditional probability of error.

2. $\lambda^{(n)} = \max_i \lambda_i$ is the maximal probability of error.

3. $P_e^{(n)} = \frac{1}{M} \sum_{i=1}^{M} \lambda_i$ is the average probability of error.

Definition (Rate)**.**
The rate R of an (M, n) code is given by $R = \frac{\log M}{n}$ bits per transmission.
A rate R is called achievable for a given channel, iff there exists a sequence of $(\lceil 2^{nR} \rceil, n)$ codes such that $\lambda^{(n)} \to 0$ as $n \to \infty$. We will write $(2^{nR}, n)$ instead of $(\lceil 2^{nR} \rceil, n)$ to shorten the notation.

Theorem (Channel Coding Theorem)**.**
Let $(\mathcal{X}, p(y|x), \mathcal{Y})$ a discrete memoryless channel.
R is achievable for all $R < C$, that means: $\forall\ R < C\ \exists$ sequence of $(2^{nR}, n)$ codes s.t. $\lambda^{(n)} \to 0$.
Conversely, any sequence of $(2^{nR}, n)$codes with $\lambda^{(n)} \to 0$ has rate $R \leq C$.

Lemma (Fano's inequality)**.**
For a discrete memoryless channel with an input message W uniformly distributed over $2^{nR}$and $\widehat{W}$ the output message

$$H(W|\widehat{W}) \leq 1 + P_e^{(n)} nR.$$

Lemma.
Let Y^n be the result of X^n passing a discrete memoryless channel of capacity C. Then for all $p(x^n)$

$$I(X^n; Y^n) \leq nC.$$

References

[1] Thomas A. Cover and Joy A. Thomas, *Elements of Information theory*, Wiley-Interscience, second edition (2006)

[2] Michael Wolf, lecture notes 2012/13, `http://www-m5.ma.tum.de/foswiki/pub/M5/Allgemeines/MA5103_2012W/lecture7.pdf`

[3] Michael Wolf, lecture notes 2012/13, `http://www-m5.ma.tum.de/foswiki/pub/M5/Allgemeines/MA5103_2012W/lecture8.pdf`