Def. 4.9.

Suppose a source produces R bits per second (rate R).

Hence, NR bits in N seconds.

Total no. of messages in N seconds: 2^{NR} (ass. integer)

M codewords available for encoding all messages

$$M = 2^{NR} \Leftrightarrow R = \frac{\log M}{N}$$

(No. of bits per channel use.)

La. 4.13.

$(\underline{X}_N, \underline{Y}_N)$ is a DMC iff $\forall \ell = 1, \dots, N$

$$P(Y_\ell = b_\ell \mid X_1 = a_1, \dots, X_N = a_N, Y_1 = b_1, \dots, Y_{\ell-1} = b_{\ell-1})$$

$$= P(Y_1 = b_\ell \mid X_1 = a_\ell) \quad \lrcorner$$

Proof. "$\Leftarrow$"

$$P(\underline{Y}_N = \underline{b}_N \mid \underline{X}_N = \underline{a}_N)$$

$$= P(\underline{Y}_N = \underline{b}_N \mid \underline{X}_N = \underline{a}_N, \underline{Y}_{N-1} = \underline{b}_{N-1}) \cdot \frac{P(\underline{Y}_{N-1} = \underline{b}_{N-1}, \underline{X}_N = \underline{a}_N)}{P(\underline{X}_N = \underline{a}_N)}$$

$$\overset{(\text{Ass.})}{=} P(Y_1 = b_N \mid X_1 = a_N) \cdot P(\underline{Y}_{N-1} = \underline{b}_{N-1}, \underline{X}_N = \underline{a}_N)$$

$$= P(Y_1 = b_N \mid X_1 = a_N) \cdot P(Y_1 = b_{N-1} \mid X_1 = a_{N-1}) \cdot P(\underline{Y}_{N-2} = \underline{b}_{N-2} \mid \underline{X}_{N-2} = \underline{a}_{N-2})$$

$$= \dots = \prod_{i=1}^{N} P(Y_1 = b_i \mid X_1 = a_i)$$

"$\Rightarrow$"

$$P(Y_\ell = b_\ell \mid \underline{X}_N = \underline{a}_N, \underline{Y}_{\ell-1} = \underline{b}_{\ell-1})$$

$$= \frac{P(\underline{Y}_\ell = \underline{b}_\ell \mid \underline{X}_N = \underline{a}_N)}{P(\underline{Y}_{\ell-1} = \underline{b}_{\ell-1} \mid \underline{X}_N = \underline{a}_N)}$$

$$\neq \frac{\sum_{b_{\ell+1},\dots,b_N \in \mathcal{Y}} P(\underline{Y}_N = \underline{b}_N \mid \underline{X}_N = \underline{a}_N)}{\sum_{b_\ell,\dots,b_N \in \mathcal{Y}} P(\underline{Y}_N = \underline{b}_N \mid \underline{X}_N = \underline{a}_N)}$$

Ex.

$$\overset{\text{Ass.}}{=} P(Y_1 = b_\ell \mid X_1 = a_\ell). \qquad \blacksquare$$

$\{(X_n, Y_n)\}$ is a sequence of independent r.v.s then $(\underline{X}_N, \underline{Y}_N)$ forms a DMC.

Th. 4.14. (Outline of the proof)

Use random coding, i.e., r.v.

$C_1, \dots, C_M \in \mathcal{X}^N$, $C_i = (C_{i1}, \dots, C_{iN})$, $i = 1, \dots, M$

with $C_{ij} \in \mathcal{X}$, i.i.d. $\sim p(x)$, $i = 1, \dots, M$, $j = 1, \dots, N$.

Th. A. For a DMC with ML-decoding it holds for all $0 \le \gamma \le 1$, $j = 1, \dots, M$

$$E(e_j(C_1, \dots, C_M)) \le (M-1)^\gamma \left(\sum_{j=1}^{d} \left(\sum_{i=1}^{m} p_i \, p_1(y_j | x_i)^{\frac{1}{1+\gamma}} \right)^{1+\gamma} \right)^N$$

$\lrcorner$

Set $G(\gamma, p) = -\ln \left[\sum_{j=1}^{d} \left(\sum_{i=1}^{m} p_i \, p_1(y_j | x_i)^{\frac{1}{1+\gamma}} \right)^{1+\gamma} \right]$

and $R = \dfrac{\ln M}{N}$:

$$E(e_j(C_1, \dots, C_M)) \le \exp(-N(G(\gamma, p) - \gamma R))$$

Set $G^*(R) = \max\limits_{0 \le \gamma \le 1} \max\limits_{p} \{ G(\gamma, p) - \gamma R \}$

Th. B. For a DMC with ML decoding there exists a code $c_1, \dots, c_M \in \mathcal{X}^N$ s.t.

$$\hat{e}(c_1, \dots, c_M) \le 4 e^{-N G^*(R)}$$

$\lrcorner$

Proof. Use $2M$ random codewords. Then

$$\frac{1}{2M}\sum_{j=1}^{2M} E\left(e_j(C_1,\dots,C_{2M})\right) \le e^{-NG^*\left(\frac{\ln 2M}{N}\right)}$$

There exists a sample $c_1,\dots,c_{2M}$ s.t.

$$\frac{1}{2M}\sum_{j=1}^{2M} e_j(c_1,\dots,c_{2M}) \le e^{-NG^*\left(\frac{\ln 2M}{N}\right)} \quad (*)$$

Remove M codewords, particularly with

$$e_k(c_1,\dots,c_{2M}) > 2\, e^{-NG^*\left(\frac{\ln 2M}{N}\right)}$$

There are at most M, otherwise $(*)$ would be violated.

For the remaining ones

$$e_j(c_{i_1},\dots,c_{i_M}) \le 4\, e^{-NG^*\left(\overset{R}{\cancel{\frac{\ln 2M}{N}}}\right)} \quad \forall j=1,\dots,M. \quad \square$$

Th. C. If $R = \frac{\ln M}{N} < C$, then

$$G^*(R) = \max_p \max_{0\le\gamma\le1}\{G(\gamma,p)-\gamma R\}$$

$$\ge \max_{0\le\gamma\le1}\{G(\gamma,p^*)-\gamma R\} > 0$$

where p^* denotes the capacity-achieving distr. $\lrcorner$

(Ref. RM, p. 103-114)

RWTH AACHEN UNIVERSITY

5. Rate Distortion Theory

Motivation:

a) By the source coding theorem (Th. 3.7 & 3.9): error free / lossless encoding needs at least on average $H(X)$ bits per symbol. What can be said if fewer bits are available.

b) Signal is represented by bits. What is the min. no. of bits needed not to exceed a certain max. distortion.

Example 5.1.

a) Representing a real number by K bits:

$$\mathcal{X} = \mathbb{R}, \quad \hat{\mathcal{X}} = \{(b_1, \dots, b_K) \mid b_i \in \{0,1\}\}$$

b) 1-bit quantization

$$\mathcal{X} = \mathbb{R}, \quad \hat{\mathcal{X}} = \{0,1\}$$

c) Representing a 28x28 gray-scale picture (8 values) by R bits

$$\mathcal{X} = 2^{28 \cdot 28 \cdot 3}, \quad \hat{\mathcal{X}} = \{1, 2, \dots, 2^R\}$$

$\mathcal{X}$ is called source alphabet, $\hat{\mathcal{X}}$ reproduction alph.. Both are assumed to be finite.

Thinking the Future
Zukunft denken

General situation:

$X_1, \dots, X_n$ i.i.d. r.v. $\sim p(x)$, $x \in \mathcal{X}$ output of some source.

$f_n(X^n)$ encoding of X^n by an index $1, 2, \dots, 2^{nR}$.

$g_n : \{1, \dots, 2^{nR}\} \to \hat{\mathcal{X}}$ decoding by a reproduction in $\hat{\mathcal{X}}$.

$$X^n \to \boxed{\begin{array}{c}\text{encoder}\\ f_n\end{array}} \xrightarrow{f_n(X^n) \in \{1, \dots, 2^{nR}\}} \boxed{\begin{array}{c}\text{decoder}\\ g_n\end{array}} \to \hat{X}^n$$

Def. 5.2. A distortion function/measure is a mapping $d : \mathcal{X} \times \hat{\mathcal{X}} \to \mathbb{R}_+$.

Examples.

a) Hamming distance, $\mathcal{X} = \hat{\mathcal{X}} = \{0, 1\}$

$$d(x, \hat{x}) = \begin{cases} 0, & x = \hat{x} \\ 1, & \text{otherwise} \end{cases}$$

b) Squared error: $d(\hat{x}, x) = (x - \hat{x})^2$

Def. 5.3. The distortion measure between sequences $x^n, \hat{x}^n$ is defined

$$d(x^n, \hat{x}^n) = \frac{1}{n} \sum_{i=1}^{n} d(x_i, \hat{x}_i).$$