**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)

$M$ code words available for encoding all messages

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

(No. of bits per channel use.)

**La. 4.13.**

$(X_N, Y_N)$ is a DMC iff $\forall \ell = 1, \ldots, N$

$$P(Y_{\ell} = b_{\ell} \mid X_1 = a_1, \ldots, X_{\ell-1} = a_{\ell-1}, Y_1 = b_1, \ldots, Y_{\ell-1} = b_{\ell-1})$$

$$= P(Y_1 = b_1 \mid X_1 = a_1)$$

**Proof.** $\Leftarrow$

$$P(Y_N = b_N \mid X_N = a_N)$$

$$= P(Y_{N-1} = b_{N-1} \mid X_{N-1} = a_{N-1}, Y_{N-1} = b_{N-1}, X_N = a_N) \cdot \frac{P(X_{N-1} = b_{N-1}, X_N = a_N)}{P(X_N = a_N)}$$

(Ass.)

$$= P(Y_1 = b_1 \mid X_1 = a_1) \cdot P(Y_{N-1} = b_{N-1} \mid X_{N-1} = a_{N-1}, Y_{N-1} = b_{N-1})$$

$$= \cdots = \frac{1}{N} \cdot P(Y_1 = b_1 \mid X_1 = a_1)$$

Thinking the Future
Zukunft denken

-1-
\[ P(Y_e = y_e \mid X_N = a_N, Y_{e-1} = y_{e-1}) = \frac{P(Y_e = y_e \mid X_N = a_N)}{P(Y_{e-1} = y_{e-1} \mid X_N = a_N)} \]

\[ \sum_{b_{e+1} \cdots b_N \in \mathcal{Y}} \frac{P(Y_N = b_N \mid X_N = a_N) \cdot P(Y_N = b_N \mid X_N = a_N)}{P(Y_1 = b_1 \mid X_1 = a_N)} \]

Assume \( P(Y_1 = b_1 \mid X_1 = a_N) \).

\[
\{ (X_N, Y_N) \} \text{ is a sequence of independent r.v.s then } (X_N, Y_N) \text{ forms a DMC.}
\]
Thinking the Future
Zukunft denken

Th. 8. If exists a code $C \subseteq \mathcal{X}^N$ with ML-decoding, then

$$R = \frac{1}{N} \max_{(c_i)_{i=1}^{\infty}} \sum_{i=1}^{\infty} I(c_i;C_{i+1})$$

Set $G(c) = -\frac{1}{N} \sum_{i=1}^{\infty} \log \frac{P(c_{i+1}|c_i)}{P(c_{i+1})}$ and

$$E(c) = \sum_{(c_i)_{i=1}^{\infty}} P_c(c) G(c)$$

For all $G$, we define $C_i = (C_{i+j}^{i+j-1})_{j=1}^{N}$ with $C_1 \in \mathcal{X}$.

Hence, for a DMC with ML-decoding, it holds

$$C_1, \ldots, C_N \in \mathcal{X} \Rightarrow \min \{E(c) : c \in \mathcal{X}^N\} = 0$$

Outline of the proof

Use standard coding, i.e., $N$.
Proof. Use $2M$ random codewords. Then

$$\frac{1}{2M} \sum_{j=1}^{2M} E(e_j(C_{2M})) \leq e^{-NG^*(\frac{b_2(M)}{N})}$$

There exists a sample $C_{1,\ldots,2M}$ s.t.

$$\frac{1}{2M} \sum_{j=1}^{2M} e_j(C_{1,\ldots,2M}) \leq e^{-NG^*(\frac{b_2(M)}{N})} \quad (\ast)$$

Remove $M$ codewords, particularly with

$$e_R(C_{1,\ldots,2M}) \geq 2e^{-NG^*(\frac{b_2(M)}{N})}$$

There are at most $M$, otherwise (\ast) would be violated. For the remaining ones

$$e_j(C_{1,\ldots,2M}) \leq 4e^{-NG^*(\frac{b_2(M)}{N})} \quad \forall j=1,\ldots,M.$$  \[ \Box \]

Th. C. If $R = \frac{b_2(M)}{N} < C$, then

$$G(R) = \max_{p} \max_{\alpha, y \in 1} \left\{ G(y; p) - \alpha R \right\}$$

$$\geq \max_{\alpha, y \in 1} \left\{ G(y; p^*) - \alpha R \right\} > 0$$

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

(Ref. RM, p. 103-114)
5. Rate Distortion Theory

Motivation:

a) By the source coding theorem (Ch. 3.22 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:
$$X = \mathbb{R}, \quad \mathcal{X} = \{b_1, \ldots, b_K \mid b_i \in \{0,1\}\}$$

b) 1-bit quantization
$$X = \mathbb{R}, \quad \mathcal{X} = \{0,1\}$$

c) Representing a $28 \times 28$ grayscale picture (8 values) by $R$ bits
$$X = 2^{28 \cdot 28 \cdot 3}, \quad \mathcal{X} = \{1,2,\ldots, 2^R\}$$

$\mathcal{X}$ is called source alphabet, $\mathcal{X}$ reproduction alphabet.
Both are assumed to be finite.
General situation:
\[ x_1, \ldots, x_n \text{ i.i.d. } n.v. ~ p(x), x \in \mathbb{F} \text{ output of some source.} \]

\( f_n(X^n) \) encoding of \( X^n \) by an index \( 1, 2, \ldots, 2^{NR} \).
\( g_n : \{1, \ldots, 2^{NR}\} \rightarrow \mathbb{X} \) decoding by a reproduction
\( \hat{x} \).

\[ x^n \rightarrow \square \text{ encoder } f_n \rightarrow f_n(X^n) \in \{1, \ldots, 2^{NR}\} \rightarrow \square \text{ decoder } g_n \rightarrow \hat{x}^n \]

Def. 5.2. A distortion function/measure is a mapping \( d : \mathbb{X} \times \mathbb{X} \rightarrow \mathbb{R}_+ \).

Examples.
\( a) \) Hamming distance, \( \mathbb{X} = \mathbb{F} = \{0, 1\} \)
\[ d(x, \hat{x}) = \begin{cases} 0, & x = \hat{x} \\ 1, & \text{otherwise} \end{cases} \]

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

Def. 5.3. The distortion measure between sequences \( x^n, \hat{x}^n \) is achieved
\[ d(x^n, \hat{x}^n) = \frac{1}{n} \sum_{i=1}^{n} d(x_i, \hat{x}_i). \]