Recapulation:

$X^n = (X_1, \dots, X_n) \in \mathcal{X}^n$, $X_1, \dots, X_n$ i.i.d. $\sim p(x)$

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

$$\hat{X}^n = (\hat{X}_1, \dots, \hat{X}_n) \in \hat{\mathcal{X}}^n$$

Def. 5.3. Distortion measure

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

Def. 5.4. A $(2^{nR}, n)$-rate distortion code of rate R and block length n consists an encoder

$$f_n : \mathcal{X}^n \to \{1, 2, \dots, 2^{nR}\}$$

and a decoder

$$g_n : \{1, 2, \dots, 2^{nR}\} \to \hat{\mathcal{X}}^n .$$

The expected distortion of the (f_n, g_n) is

$$D = E\, d(X^n, \hat{X}^n) = E\, d(X^n, g_n(f_n(X^n))).$$ ┘

RWTH AACHEN UNIVERSITY

Remarks.

a) $\mathcal{X}, \hat{\mathcal{X}}$ are assumed to be finite.

b) 2^{nR} means $\lceil 2^{nR} \rceil$, if it is not integer.

c) f_n yields 2^{nR} different values.
We need $\approx nR$ bits to represent each. Hence,
R = no. of bits per source symbol needed to represent $f_n(X^n)$.

d)
$$D = E\,d(X^n, \hat{X}^n)$$
$$= E\,d(X^n, g_n(f_n(X^n)))$$
$$= \sum_{x^n \in \mathcal{X}^n} p(x^n)\, d(x^n, g_n(f_n(x^n)))$$

e) $\{g_n(1), \ldots, g_n(2^{nR})\}$ is called <u>codebook</u>.
$f_n^{-1}(1), \ldots, f_n^{-1}(2^{nR})$ are called <u>assignment regions</u>.

Ultimate goal of lossy source coding:
- minimize R for a given D or
- minimize D for a given R.

Thinking the Future
Zukunft denken

<u>Def. 5.5.</u> A rate distortion pair (R,D) is called <u>achievable</u> if there exists a sequence of $(2^{nR}, n)$-rate distortion codes such that

$$\lim_{n\to\infty} E\, d(X^n, g_n(f_n(X^n))) \leq D . \quad \lrcorner$$

<u>Def. 5.6.</u> The rate distortion function is defined as

$$R(D) = \inf_{(R,D) \text{ is achievable}} R \quad \lrcorner$$

<u>Def. 5.7.</u> The information ~~distr~~ rate distortion function $R_I(D)$ is defined as follows

$$R_I(D) = \min_{p(\hat{x}|x) : \sum_{(x,\hat{x})} p(x,\hat{x})\, d(x,\hat{x}) \leq D} I(X,\hat{X})$$

$$= \min_{p(\hat{x}|x) : E\, d(X,\hat{X}) \leq D} I(X,\hat{X}) . \quad \lrcorner$$

Compare with capacity:

C : given $p(\hat{x}|x)$, max $I(X,\hat{X})$ over the input distr

$R_I(D)$: given $p(x)$, min $I(X,\hat{X})$ over "channels" s.t. the expected distortion does exceed D.

RWTH AACHEN UNIVERSITY

Theorem 5.8.

a) $R_I(D)$ is a convex nonincreasing fct. of D.

b) $R_I(D) = 0$, if $D > D^*$, $D^* = \min_{\hat{x} \in \hat{\mathcal{X}}} E\, d(X, \hat{x})$

c) $R_I(0) \le H(X)$ ┘

Proof. Yeung, p. 198 ff, Cover & Thomas p. 316 ff.

Theorem 5.9. $X \sim Bin(1, p)$, $P(X=0) = 1-p$, $P(X=1) = p$, $0 \le p \le 1$,

d : Hamming distance.

$$R_I(D) = \begin{cases} H(p) - H(D), & 0 \le D \le \min\{p, 1-p\} \\ 0, & \text{otherwise} \end{cases}$$

Proof. W.l.o.g. assume $p < \frac{1}{2}$, otherwise interchange 0 and 1.

$$R_I(D) = \min_{p(\hat{x}|x) : E\, d(X, \hat{X}) \le D} I(X; \hat{X})$$

Assume $D \le p < \frac{1}{2}$

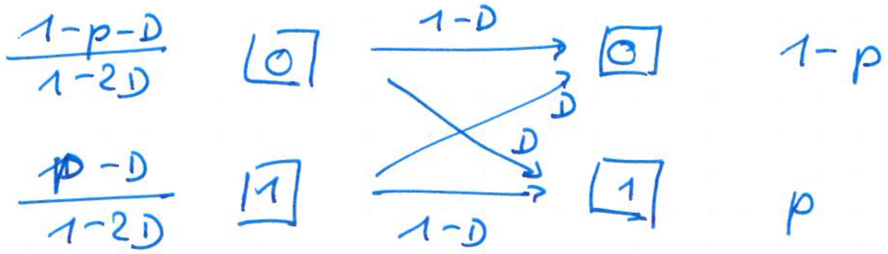
$$\begin{aligned} I(X; \hat{X}) &= H(X) - H(X | \hat{X}) \\ &= H(X) - H(X \oplus \hat{X} | \hat{X}) && \left[X \oplus \hat{X} \text{ given } \hat{X} \text{ has the same entropy as } X \text{ given } \hat{X} \right. \\ &\ge H(X) - H(X \oplus \hat{X}) \\ &= H(p) - H(P(X \ne \hat{X})) && \left[P(X \oplus \hat{X} = 1) = P(X \ne \hat{X}) \right. \\ &\ge H(p) - H(D) && \left[P(X \ne \hat{X}) = E\, d(X, \hat{X}) \le D < \tfrac{1}{2} \right. \end{aligned}$$

Thinking the Future
Zukunft denken

This lower bound is attained by the following joint distribution of $(X, \hat{X})$

$X \backslash \hat{X}$	0	1	
0	$\frac{(1-D)(1-p-D)}{1-2D}$	$\frac{D(p-D)}{1-2D}$	$1-p$
1	$\frac{D(1-p-D)}{1-2D}$	$\frac{1-D(p-D)}{1-2D}$	p
	$\frac{1-p-D}{1-2D}$	$\frac{p-D}{1-2D}$	1

Corresponds to a BSC



It follows that

$$P(X \neq \hat{X}) = E\,d(X, \hat{X})$$
$$= \frac{D(p-D)}{1-2D} + \frac{D(1-p-D)}{1-2D} = D$$

Further

$$I(X;\hat{X}) = H(X) - H(X|\hat{X})$$
$$= H(p) - \Big[\underbrace{H(X|\hat{X}=0)}_{=H(D)} P(\hat{X}=0) + \underbrace{H(X|\hat{X}=1)}_{=H(1-D)=H(D)} P(\hat{X}=1)\Big]$$
$$= H(p) - H(D)$$

such that the lower bound is attained.

If $D \geq p$ set $P(\hat{X}=0)=1$ and set

$X \backslash \hat{X}$	0	1	
0	$1-p$	0	$1-p$
1	p	0	p
	1	0	

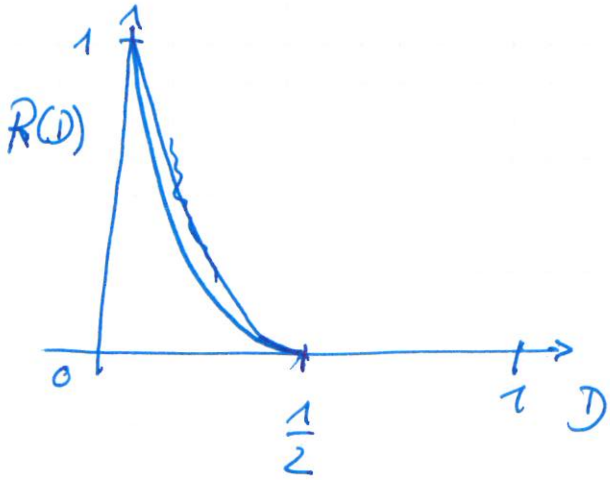
Then $E\,d(X,\hat{X}) = P(X \neq \hat{X}) = P(X=1) = p \leq D$

and $I(X;\hat{X}) = H(X) - H(X|\hat{X})$

$$= H(p) - H(X|\hat{X}=0)\cdot 1$$

$$= H(p) - H(p) = 0$$

Plot for $Bin(1, \frac{1}{2})$:



RWTH AACHEN UNIVERSITY

Theorem 5.10. (Converse to the rate distortion theorem)

$$R(D) \geq R_I(D)$$

Proof.

Recall the general situation: $X_1, \dots, X_n$ i.i.d. $\sim X \sim p(x), x \in \mathcal{X}$.

$\hat{X}^n = g_n(f_n(X^n))$ has at most 2^{nR} values. Hence

$$H(\hat{X}^n) \leq \log 2^{nR} \leq nR$$

We first show: (R,D) achievable $\Rightarrow R \geq R_I(D)$

Suppose (R,D) is achievable. Then

$$
\begin{aligned}
nR &\geq H(\hat{X}^n) \\
&\geq H(\hat{X}^n) - H(\hat{X}^n | X^n) \\
&= I(\hat{X}^n; X^n) = I(X^n, \hat{X}^n) \\
&= H(X^n) - H(X^n | \hat{X}^n) \\
&= \sum_{i=1}^n H(X_i) - \sum_{i=1}^n H(X_i | \hat{X}^n, (X_1, \dots, X_{i-1})) \\
&\geq \sum_{i=1}^n I(X_i; \hat{X}_i) \\
&\geq \sum_{i=1}^n R_I(Ed(X_i, \hat{X}_i)) \\
&= n \sum_{i=1}^n \frac{1}{n} R_I(Ed(X_i, \hat{X}_i)) \\
&\geq n R_I\Big(\frac{1}{n} \sum_{i=1}^n Ed(X_i, \hat{X}_i)\Big) \qquad (R_I \text{ convex}) \\
&= n R_I\, Ed(X^n, \hat{X}^n) \qquad Ed(X^n, \hat{X}^n) \leq D \\
&\geq n R_I(D) \qquad (R_I \text{ non-increasing})
\end{aligned}
$$

Thinking the Future
Zukunft denken

Hence, $\inf_{(R,D) \text{ achievable}} R = R(D) \geq R_I(D)$ ∎

The reverse inequality in Th. 5.10. also holds.

Th. 5.11. (The Rate Distortion Theorem)

$$R(D) = R_I(D). \quad \lrcorner$$

Proof. ~~Def~~ $R(D) \geq R_I(D)$ ✓ Th. 5.10.

$R_I(D) \leq R(D)$

Yeung: Section 9.5, p. 206-212

Cover & Thomas: Section 10.5, p. 318-324 $\lrcorner$