Noisy Channel Coding:

Correlated Random Variables & Communication over a Noisy Channel

Toni Hirvonen

Helsinki University of Technology

Laboratory of Acoustics and Audio Signal Processing

Toni.Hirvonen@hut.fi

T-61.182 Special Course in Information Science II / Spring 2004

Contents

- More entropy definitions
  - joint & conditional entropy
  - mutual information
- Communication over a noisy channel
  - overview
  - information conveyed by a channel
  - noisy channel coding theorem

Joint Entropy

Joint entropy of X, Y is:

$$H(X,Y) = \sum_{xy \in \mathcal{A}_X \mathcal{A}_Y} P(x,y) \log \frac{1}{P(x,y)}$$

Entropy is additive for independent random variables:

$$H(X,Y) = H(X) + H(Y) \text{ iff } P(x,y) = P(x)P(y)$$

Conditional Entropy

Conditional entropy of X given Y is:

$$H(X|Y) = \sum_{y \in \mathcal{A}_Y} P(y) \left[\sum_{x \in \mathcal{A}_X} P(x|y) \log \frac{1}{P(x|y)} \right] = \sum_{y \in \mathcal{A}_X \mathcal{A}_Y} P(x,y) \log \frac{1}{P(x|y)}$$

It measures the average uncertainty (*i.e.* information content) that remains about x when y is known.

Mutual Information

Mutual information between X and Y is:

$$I(Y;X) = I(X;Y) = H(X) - H(X|Y) \geq 0$$

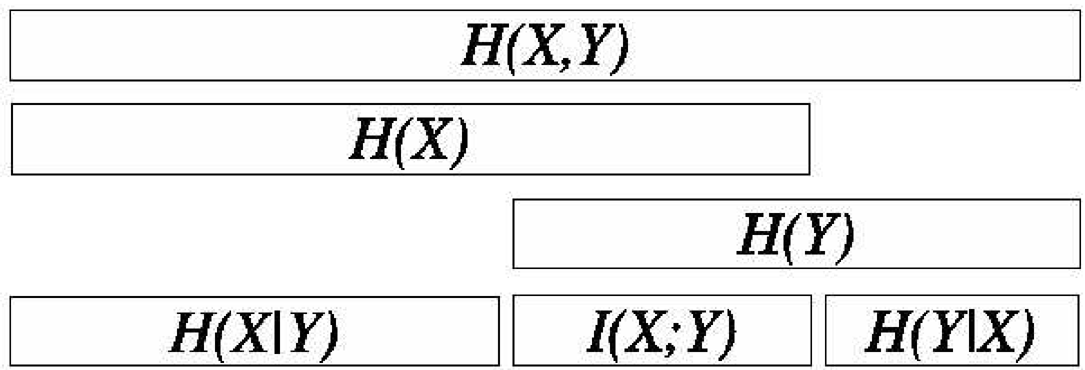
It measures the average reduction in uncertainty about x that results from learning the value of y, or vice versa.

Conditional mutual information between X and Y given Z is:

$$I(Y;X|Z) = H(X|Z) - H(X|Y,Z)$$

Breakdown of Entropy

Entropy relations:



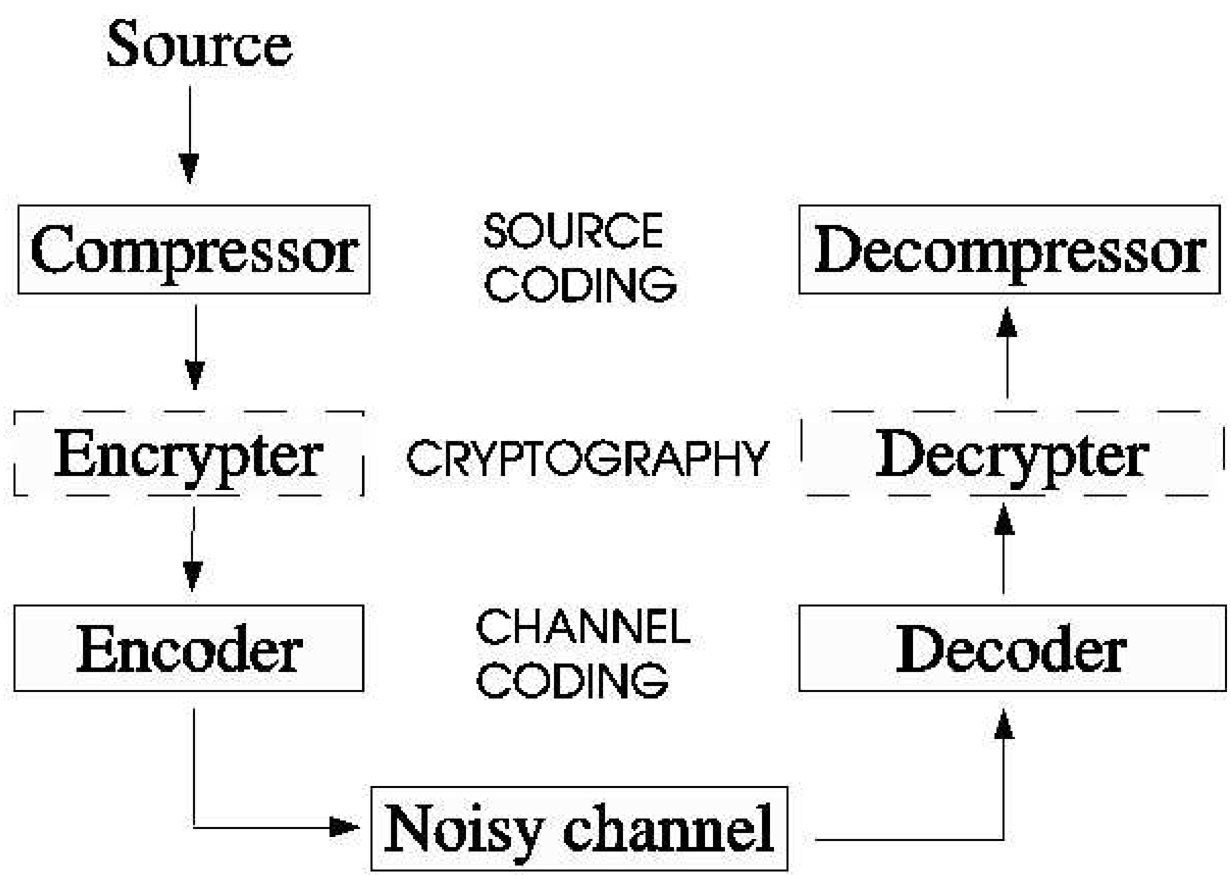
Chain rule of entropy:

$$H(X,Y) = H(X) + H(Y|X) = H(Y) + H(X|Y)$$

Noisy Channel: Overview

- Real-life communication channels are hopelessly noisy *i.e.* introduce transmission errors
- However, a solution can be achieved
  - the aim of source coding is to remove redundancy from the source data
  - the aim of channel coding is to make a noisy channel behave like a noiseless one via controlled adding of redundancy

Noisy Channel: Overview (Cont.)



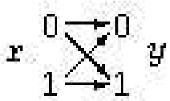
Noisy Channels

- General discrete memoryless channel is characterized by:
  - input alphabet $\mathcal{A}_X$
  - output alphabet $\mathcal{A}_Y$
  - set of conditional probability distributions $P(y|x)$, one for each $x \in \mathcal{A}_X$
- These transition probabilities can be written in a matrix form:

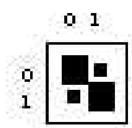
$$Q_{j|i} = P(y = b_j | x = a_i)$$

Noisy Channels: Useful Models

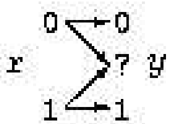
Binary symmetric channel. $\mathcal{A}_X = \{0,1\}$. $\mathcal{A}_Y = \{0,1\}$.



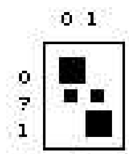
$$\begin{array}{rclrcl} P(y=0 \mid x=0) & = & 1-f; & P(y=0 \mid x=1) & = & f; \\ P(y=1 \mid x=0) & = & f; & P(y=1 \mid x=1) & = & 1-f. \end{array}$$



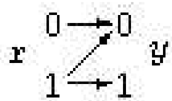
Binary erasure channel. $\mathcal{A}_X = \{0,1\}$. $\mathcal{A}_Y = \{0,?,1\}$.



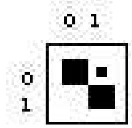
$$\begin{array}{rclrcl} P(y=0 \mid x=0) & = & 1-f; & P(y=0 \mid x=1) & = & 0; \\ P(y=? \mid x=0) & = & f; & P(y=? \mid x=1) & = & f; \\ P(y=1 \mid x=0) & = & 0; & P(y=1 \mid x=1) & = & 1-f. \end{array}$$



Z channel. $\mathcal{A}_X = \{0,1\}$. $\mathcal{A}_Y = \{0,1\}$.



$$\begin{array}{rclrcl} P(y=0 \mid x=0) & = & 1; & P(y=0 \mid x=1) & = & f; \\ P(y=1 \mid x=0) & = & 0; & P(y=1 \mid x=1) & = & 1-f. \end{array}$$



Inferring Channel Input

- If we receive symbol y, what is the probability of input symbol x?
- Let's use the Bayes' theorem:

$$P(x|y) = \frac{P(y|x)P(x)}{P(y)} = \frac{P(y|x)P(x)}{\sum_{x'} P(y|x')P(x')}$$

Example: a Z-channel has $f = 0.15$ and the input probabilities (*i.e.* ensemble) $p(x = 0) = 0.9, p(x = 1) = 0.1$. If we observe $y = 0$,

$$P(x = 1|y = 0)) = \frac{0.15 * 0.1}{0.15 * 0.1 + 1 * 0.9} = 0.26$$

Information Transmission over a Channel

- What is a suitable measure for transmitted information?
- Given what we know, the mutual information $I(X;Y)$ between the source X and the received signal Y is sufficient
  - remember that:
    $I(Y;X) = I(X;Y) = H(X) - H(X|Y)$
    $=$ the average reduction in uncertainty about x that results from learning the value of y, or vice versa.
  - on average, y conveys information about x if $H(X|Y) < H(X)$

Information Transmission over a Channel (Cont.)

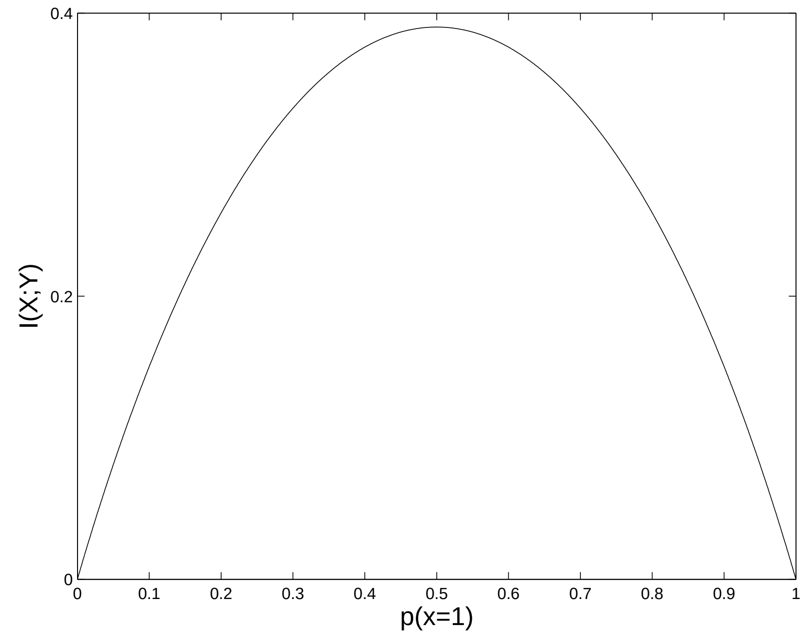
- In real life, we are interested in communicating over a channel with a negligible probability of error
- How can we combine this idea with the mathematical expression of conveyed information, it i.e.
  $I(X;Y) = H(X) - H(X|Y)$
- Often it is more convenient to calculate mutual information as
  $I(X;Y) = H(Y) - H(Y|X)$

Information Transmission over a Channel (Cont.)

- Mutual information between the input and the output depends on the input ensemble $\mathcal{P}_X$
- Channel capacity is defined as the maximum of its mutual information
- The optimal input distribution maximizes mutual information

$$C(Q) = \max_{\mathcal{P}_X} I(X;Y)$$

Binary Symmetric Channel Mutual Information



$I(X;Y)$ for a binary symmetric channel with $f = 0.15$ as a function of input distribution

Noisy Channel Coding Theorem

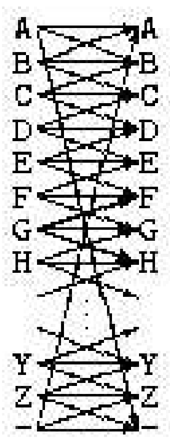
- It seems plausible that channel capacity C can be used as a measure of information conveyed by a channel
- What is not so obvious:

  Shannon's noisy channel coding theorem (pt.1):

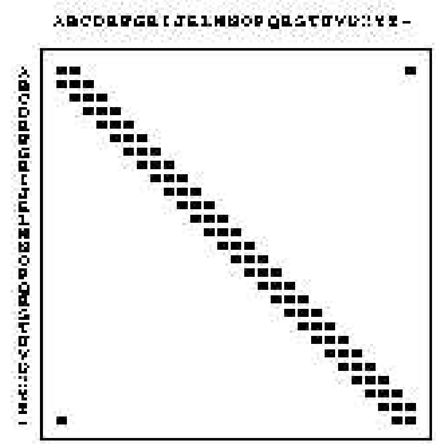
  All discrete memoryless channels have non-negative capacity C. For any $\epsilon > 0$ and $R < C$, for large enough N, there exists a block code of length N and rate $\geq R$ and a decoding algorithm, such that the maximal probability of block error is $< \epsilon$

Proving the Noisy Channel Coding Theorem

Let's consider Shannon's theorem and a *noisy typewriter* channel:



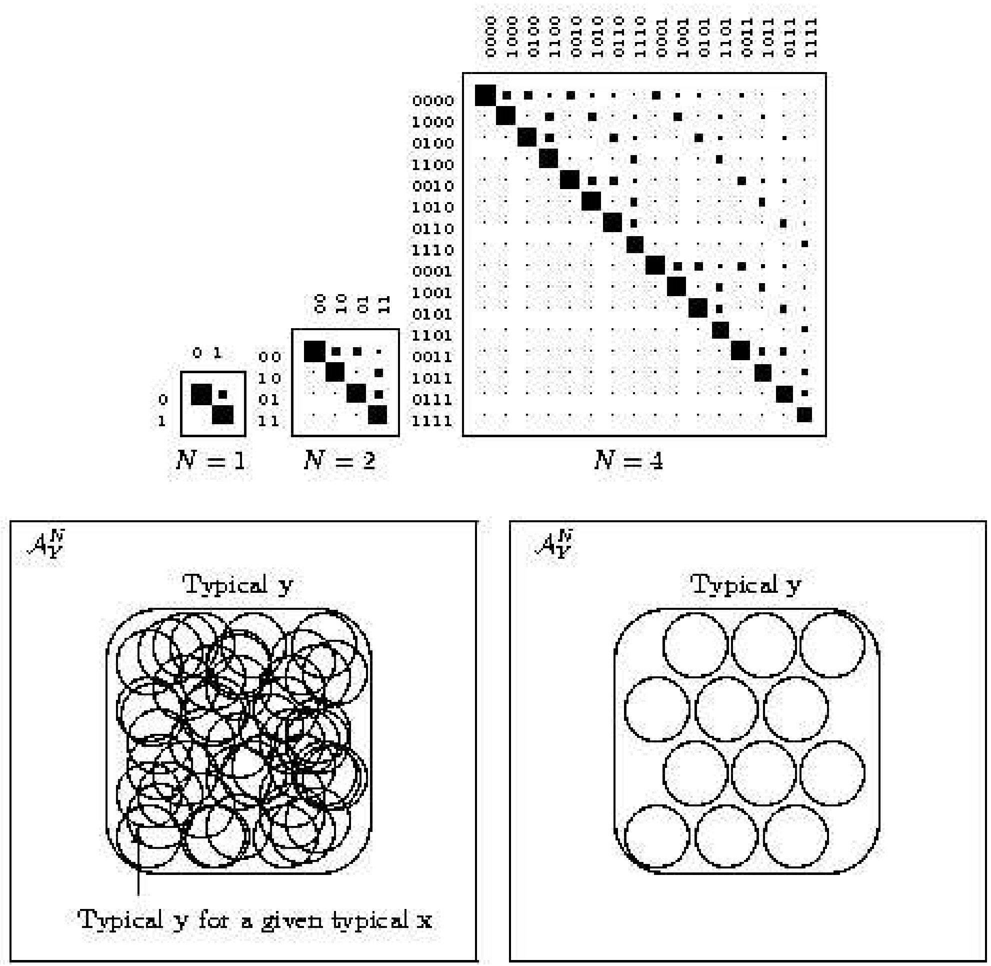
$$
\begin{aligned}
&\vdots\\
P(y=\mathrm{F} \mid x=\mathrm{G}) &= 1/3;\\
P(y=\mathrm{G} \mid x=\mathrm{G}) &= 1/3;\\
P(y=\mathrm{H} \mid x=\mathrm{G}) &= 1/3;\\
&\vdots
\end{aligned}
$$



Proving the Noisy Channel Coding Theorem (Cont.)

- Consider next *extended channels*:
  - corresponds to N uses of a single channel (block codes)
  - an extended channel has $|\mathcal{A}_x|^N$ possible inputs x and $|\mathcal{A}_y|^N$ possible outputs
- If N is large, x is likely to produce outputs only in a small subset of the output alphabet
  - extended channel looks a lot like a noisy typewriter

Example: an Extended Z-channel



Homework

- 8.10: mutual information
- 9.17: channel capacity