

S-72.2410 Information Theory (5 cr) P

Lectures: Mondays 9–12, room I346 and Wednesdays 9–12, room E110

Teacher: Prof. Patric Östergård, <http://users.tkk.fi/~pat/>
(office hour: Tuesdays 10–11, room I447B)

Tutorials: Tuesdays 14–16, hall S3 and Fridays 12–14, hall S2, first tutorial: 6.11.2007

Assistant: Vesa Vaskelainen, M.Sc. (Tech.), room I436, tel. 451 2401, firstname.lastname@tkk.fi

Home page: <http://t11.tkk.fi/en/Studies/S-72.2410>

Registration: In WWWTopi

© Patric Östergård

Closely Related Courses

Prerequisites: Basic courses in mathematics.

Channel coding:

S-72.3280 Advanced Radio Transmission Methods

S-72.3320 Advanced Digital Communication

S-72.3410 Coding Methods

Source coding:

S-88.4205 Image and Video Compression

S-89.3630 Speech Transmission Technology

Cryptography:

T-79.4501 Cryptography and Data Security

T-79.5501 Cryptology

© Patric Östergård

Learning Objectives

Upon completion of the course, the students will be able to

- ▷ define and apply the basic concepts of information theory (entropy, etc.);
- ▷ differentiate between lossy and lossless data compression methods, and describe the most common such methods;
- ▷ design an efficient data compression scheme for a given information source;
- ▷ calculate the capacity of communication channels;
- ▷ sketch Shannon's proof regarding the limits of error-free communication; and
- ▷ explain the impact of feedback and/or many senders or receivers on the communication problem.

© Patric Östergård

Literature (1)

[Cov] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, Wiley, New York, 1991. [**Course literature.**]

D. J. C. MacKay, *Information Theory, Inference, and Learning Algorithms*, Cambridge University Press, Cambridge, 2003.

R. G. Gallager, *Information Theory and Reliable Communication*, Wiley, New York, 1968.

C. E. Shannon and W. Weaver, *Mathematical Theory of Communication*, University of Illinois Press, 1963. [Written by the founder of information theory, Claude Shannon.]

© Patric Östergård

Literature (2)

J. R. Pierce, *An Introduction to Information Theory: Symbols, Signals and Noise*, 2nd ed., Dover, New York, 1980. [For non-mathematicians.]

J. Ekberg and S. J. Halme, *Informaatioteoria*, 2nd ed., Otakustantamo (Vol. 805), Espoo, 1978. [In Finnish.]

Main journal: *IEEE Transactions on Information Theory*.

Main conferences: *IEEE Symposium on Information Theory*, *IEEE Workshop on Information Theory*.

© Patric Östergård

Information Theory

Information theory answers two fundamental questions in communication theory:

Q: What is the ultimate data compression?

A: The entropy H .

Q: What is the ultimate transmission rate of communication?

A: The channel capacity C .

There is also a wide variety of applications in other fields, as shown by [Cov, Fig. 1.1].

© Patric Östergård

(Preliminary) Outline of the Course

1. Introduction (1)
2. Basic concepts in information theory (2)
3. Stochastic processes (1)
4. Data compression (2)
5. Channel capacity (2)
6. Continuous channels (1)
7. Universal coding (1)
8. Network information theory & Applications (1)
9. Project Review (1)

To pass the course: Project (P), five home assignments (A), exam (E).

Mark: $\max(E, P-1, A-2)$. It is also required that $\min(E, P, A) \geq 1$.

© Patric Östergård

Entropy

The **entropy** of a random variable X with a probability mass function $p(x)$ is defined by

$$H(X) = - \sum_x p(x) \log_2 p(x).$$

The entropy is measured in bits and is a measure of the average uncertainty in the random variable. It is the number of bits on the average required to describe the random variable.

We write $\log x := \log_2 x$ in the sequel.

© Patric Östergård

Example: Variable with Uniform Distribution

Consider a random variable with uniform distribution over 32 ($= 2^5$) outcomes. Obviously, 5-bit strings suffice to identify an outcome. The entropy is

$$H(X) = - \sum_{i=1}^{32} p(i) \log p(i) = - \sum_{i=1}^{32} \frac{1}{32} \log \frac{1}{32} = \log 32 = 5 \text{ bits,}$$

which agrees with the number of bits needed to describe X .

© Patric Östergård

Example: Variable with Nonuniform

Distribution (2)

As the win probabilities are not uniform, it makes sense to use shorter descriptions for the more probable horses, and longer descriptions for the less probable ones. For example, the following strings can be used to represent the eight horses:

0, 10, 110, 1110, 111100, 111101, 111110, 111111.

The average description length is then 2 bits (=entropy).

▷ The entropy gives a lower bound for the average description length.

© Patric Östergård

Example: Variable with Nonuniform

Distribution (1)

Assume that the probabilities of winning for eight horses taking part in a horse race are $\{\frac{1}{2}, \frac{1}{4}, \frac{1}{8}, \frac{1}{16}, \frac{1}{64}, \frac{1}{64}, \frac{1}{64}, \frac{1}{64}\}$. The entropy of this distribution (that is, of the horse race) is then

$$H(X) = -\frac{1}{2} \log \frac{1}{2} - \frac{1}{4} \log \frac{1}{4} - \frac{1}{8} \log \frac{1}{8} - \frac{1}{16} \log \frac{1}{16} - 4 \frac{1}{64} \log \frac{1}{64} = 2 \text{ bits.}$$

To send a message indicating the winner of the race, one can send the index of the winning horse (000, ..., 111); this requires 3 bits for any of the horses. But there is another (better) description.

© Patric Östergård

Mutual Information

entropy Uncertainty of a single random variable.

conditional entropy The entropy of a random variable, given another random variable.

mutual information The reduction in uncertainty due to another random variable.

For two random variables, X and Y , the mutual information is

$$I(X; Y) = H(X) - H(X|Y) = \sum_{x,y} p(x,y) \log \frac{p(x,y)}{p(x)p(y)}.$$

The mutual information is symmetric in X and Y and non-negative.

© Patric Östergård

Channel Capacity

A communication channel is a system in which the output depends probabilistically on its input. A probability transition matrix determines the conditional distribution of the output given the input. With input X and output Y , the **channel capacity** is defined by

$$C = \max_{p(x)} I(X; Y).$$

- ▷ The capacity is the maximum rate at which we can send information over the channel and recover the information at the output with a vanishingly low probability of error.

© Patric Östergård

Example: Noisy Four-Symbol Channel

In this channel, depicted in [Cov, Fig. 1.4], each input symbol is received as the same symbol with probability $1/2$ or as the (cyclically) next symbol with probability $1/2$.

If we use all symbols for this channel, a received symbol does not tell with certainty which input symbol was sent. However, if we use only two of the inputs (1 and 3, or 2 and 4), then this channel acts like the noiseless channel in the previous example. Hence, the channel capacity is at least 1 bit; in fact, it is exactly 1 bit.

© Patric Östergård

Example: Noiseless Binary Channel

For the noiseless binary channel, illustrated in [Cov, Fig. 1.3], the binary input is reproduced exactly at the output. Any transmitted bit is therefore received without error; in each transmission, we can send 1 bit reliably to the receiver and the capacity is 1 bit.

We can also calculate the capacity

$$C = \max I(X; Y) = 1 \text{ bit.}$$

© Patric Östergård

Example: Binary Symmetric Channel

The binary symmetric channel, shown in [Cov, Fig. 1.5], is the basic example of a noisy communication system.

The capacity of this channel is

$$C = 1 + p \log p + (1 - p) \log(1 - p)$$

bits per transmission. However, it is no longer obvious how one can achieve this capacity. The key is that if we use the channel many times, we get a situation similar to that in the four-symbol channel of the previous example, and we can indeed send information at a rate C bits per transmission with an arbitrarily low probability of error.

© Patric Östergård