Introduction

Mário S. Alvim
(msalvim@dcc.ufmg.br)

Information Theory

DCC-UFMG
(2017/02)

What is information: An introductory example

- Imagine Alice has an idea that she wants to share with Bob.

  Alice can convey her idea in several ways:

  1. say it out loud so Bob can hear it,
  2. write it on a piece of paper,
  3. paint it on a cave wall,
  4. write a song and sing it to Bob,
  5. light a fire and send smoke signals,
  6. ...

- In one way or another, she may convey her idea to Bob.

  But, given so many options to do so...

  1. What makes the methods different?
  2. What makes them the same?

What is information: An introductory example

- In each method of transmitting the message, Alice must use a specific **language**:

  1. the English spoken language, made of sound vibrations,
  2. the English written language, made of special symbols drawn in a specific pattern on a surface,
  3. the language of body signals,
  4. the language of smoke signals,
  5. ...

What is information: An introductory example

- Whichever method employed, we consider it to be successful if Bob becomes aware of Alice's idea, that is, if Alice transmits information to Bob.

  Information is, intuitively, related to how the state of an object (e.g., a mind) influences the state of another (e.g., another mind).

- If each method of transmitting the message uses a <u>different language</u>, but if Bob becomes aware of Alice's exactly same idea in each method, then, intuitively, the transmitted <u>information is the same</u>.

- In this context, two important questions are:

  1. How to measure the information transmitted in each method?
  2. How to most efficiently use the language, and means, available to transmit information?

  Answering these two questions are two central goals of **Information Theory**.

Information theory in a nutshell

- The first question we mentioned as a concern of Information Theory was

  "How to measure the information transmitted in each method?"

- To answer this question, Information Theory defines defines a **measure of information**, and a scale on which measure the information contents of a message.

- Intuitively:

  - **Entropy** is a scale on which we measure the information contents of a message.

    Entropy is to information what a scale is to mass: a scale is a tool to measure which one of two rocks has the biggest mass.

  - The **bit** is a universal unit to measure information.

    A bit to information what the gram is to mass: a standard unit of comparison.

Information theory in a nutshell

- The second question we mentioned as a concern of Information Theory was

  "How to most efficiently use the language, and means, available to transmit information?"

- To answer this question, Information Theory formalizes the concept of a **communication channel**, which encompasses the definition of the language used and the means by which the message is transmitted.

  In our example, communication channels available are the air (for spoken language), pieces of paper (for written language), electromagnetic waves (for radio transmission), etc.

- Information theory, then, looks for efficient ways to **encode information** to transmit it over the available communication channel, so to maximize the **transmission rate** (i.e., the number of bits transmitted per use of the channel), given the **channel capacity** (i.e., the maximum transmission rate the channel allows without introducing errors in the message).

The birth of information theory: communication theory

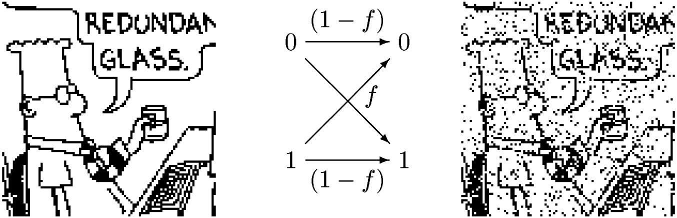
- Information theory was born in 1 948, with the publication of Claude E. Shannon's seminal paper *"A Mathematical Theory of Communication"* [link].
- According to Shannon:
  "The fundamental problem of communication is that of reproducing at one point either exactly or approximately a message selected at another point."

The problem of communication over an imperfect channel

- Consider we want to transmit an image in the following scenario.

    1. We encode the image as a matrix of monochromatic pixels: 0 for white and 1 for black.

    2. We send each pixel through a **noisy channel**, which inverts a fraction f of the bits, and delivers a fraction $1 - f$ correctly.



- For instance, the channel could be a hard disk (HD) with an error rate of f: we want to save the image in the HD and be able to reliably recover it at a later time.

The problem of communication over an imperfect channel

- There are two main approaches to solve this problem:

  1. In the **"physical" approach** we improve the hardware with more technology to reduce the error rate f until we reach a level we consider acceptable.

     In the case of our HD, it would be acceptable if it did not flip a single bit in its lifetime (as a single flipped bit can destroy your high-quality new blockbuster movie).

     If our HD is supposed to record $1GB$ of data a day, everyday for 10 years, and no error is allowed, then we would need to improve technology to get an error rate of about

     $$f \approx 10^{-15}.$$

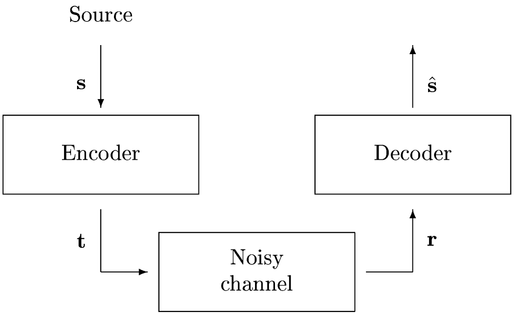
     Clearly, this is not a viable option...

The problem of communication over an imperfect channel

- There are two main approaches to solve this problem:

  2. In the **"system" approach** we accept that the error rate f of the channel is fixed, and create clever encoding/decoding schemes to create a self-correcting way of transmitting information.

     The key is to add **redundancy** to the transmitted message, so if some part of it is lost, it can be recovered.



The problem of communication over an imperfect channel

- Information theory is concerned with the viability of this system approach.
- In particular, the sub-fields of **communication theory** and **coding-theory** study the question:

  "What is the best error-correcting performance we could achieve?"

Error-correcting codes

- Error-correcting codes should:

  1. detect errors;
  2. correct the errors;
  3. do so without re-transmitting a message, i.e., there's only one chance to encode and send the message.

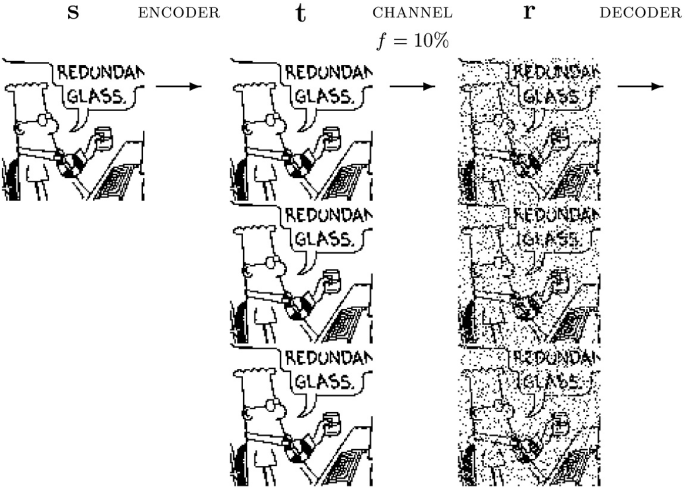
Error-correcting codes: Repetition codes

- In a repetition code R_n, each bit is repeated n times.
- R_3 code:

$$
\begin{array}{cccccccc}
\mathbf{s} & 0 & 0 & 1 & 0 & 1 & 1 & 0 \\
\mathbf{t} & \overbrace{0\,0\,0} & \overbrace{0\,0\,0} & \overbrace{1\,1\,1} & \overbrace{0\,0\,0} & \overbrace{1\,1\,1} & \overbrace{1\,1\,1} & \overbrace{0\,0\,0} \\
\mathbf{n} & 0\,0\,0 & 0\,0\,1 & 0\,0\,0 & 0\,0\,0 & 1\,0\,1 & 0\,0\,0 & 0\,0\,0 \\
\hline
\mathbf{r} & 0\,0\,0 & 0\,0\,1 & 1\,1\,1 & 0\,0\,0 & 0\,1\,0 & 1\,1\,1 & 0\,0\,0
\end{array}
$$

Error-correcting codes: Repetition codes

- Transmitting an image using an R_3 code:



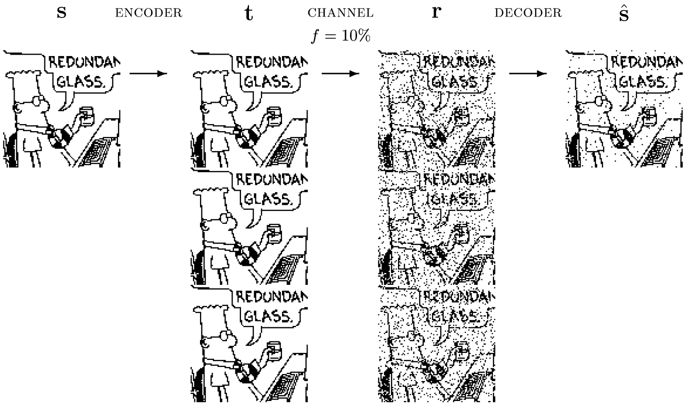
Error-correcting codes: Repetition codes

- Decoding a message using an R_3 code:

$$
\begin{array}{rccccccc}
\mathbf{s} & 0 & 0 & 1 & 0 & 1 & 1 & 0 \\
 & \overbrace{} & \overbrace{} & \overbrace{} & \overbrace{} & \overbrace{} & \overbrace{} & \overbrace{} \\
\mathbf{t} & 000 & 000 & 111 & 000 & 111 & 111 & 000 \\
\mathbf{n} & 000 & 001 & 000 & 000 & 101 & 000 & 000 \\
\hline
\mathbf{r} & \underbrace{000} & \underbrace{001} & \underbrace{111} & \underbrace{000} & \underbrace{010} & \underbrace{111} & \underbrace{000} \\
\hat{\mathbf{s}} & 0 & 0 & 1 & 0 & 0 & 1 & 0 \\
\text{corrected errors} & & \star & & & & & \\
\text{undetected errors} & & & & & \star & &
\end{array}
$$

Error-correcting codes: Repetition codes

- Recovering an image using an R_3 code:



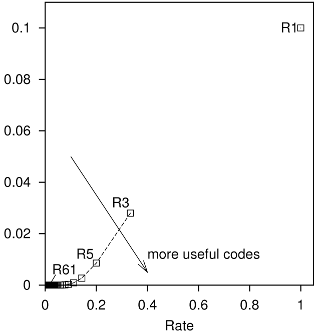
Error-correcting codes: Repetition codes

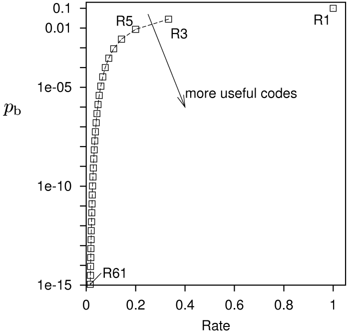
- Note that the channel we have originally:
  - has a probability of flipping a bit of $f = 0.1$, that is, the probability of a wrong bit being transmitted is $p_b = 0.1$, and
  - 1 bit of the image is transmitted in each use of the channel.
- With the R_3 code:
  - the probability of flipping a bit is reduced to $p_b \approx 0.03$, however
  - the rate channel needs to be used 3 times to transmit one bit, sot the rate of transmission was reduced to $1/3 = 0.33$ bits transmitted per use of the channel.

Error-correcting codes: Repetition codes

- We can plot the error-per-transmission (in percentage) per transmission-rate (bits/use of the channel).

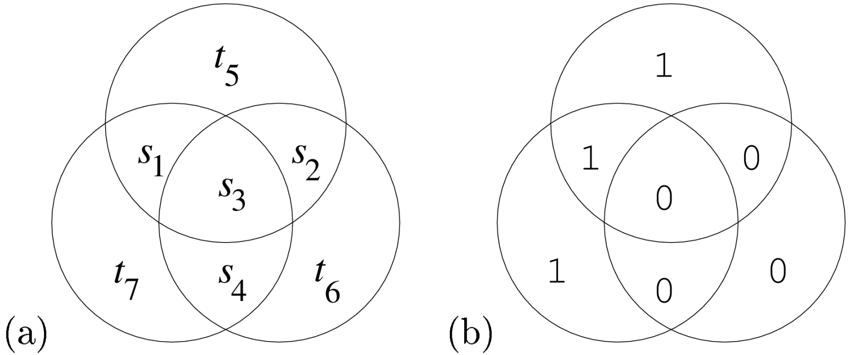
  For various R_n codes (in the second graph, the y-axis is in log-scale):





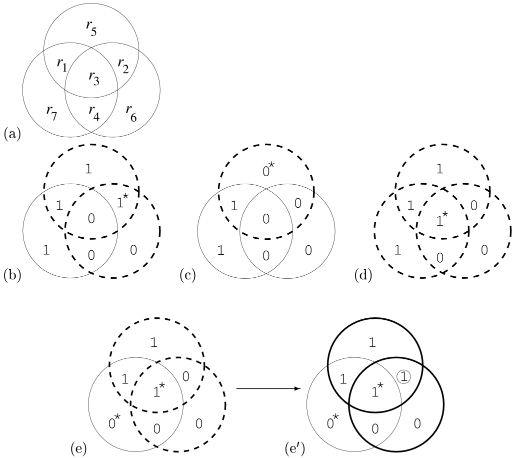
Error-correcting codes: Hamming codes

- Hamming code: encoding blocks of bits instead of individual bits, and parity bits are added.
- The $H(7,4)$ Hamming code transmits blocks of 7 bits, out of which 4 are real bits and $7-4=3$ bits are redundancy added.



Error-correcting codes: Hamming codes

- Hamming code: decoding exploits redundancy.

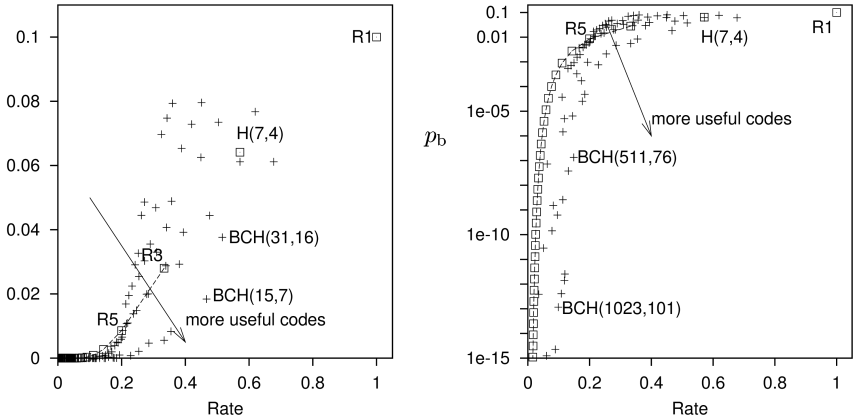


Error-correcting codes: Hamming codes

- With the $(7,4)$ Hamming code:
  - the probability of flipping a bit is reduced to $p_b \approx 0.07$, however
  - the rate channel needs to be used 7 times to transmit 4 bits, sot the rate of transmission was reduced to $4/7 \approx 0.57$ bits transmitted per use of the channel.

Error-correcting codes: comparison

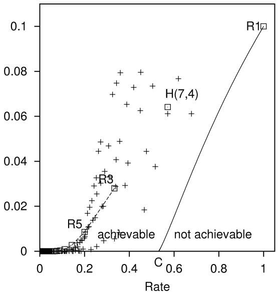
- For various R_n repetition codes and (n, r) Hamming codes:

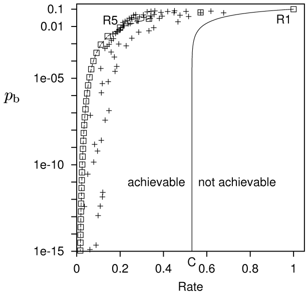


The solution for the communication problem: channel capacity

- But what is the best trade-off between error and transmission rate we can achieve?
- Shannon solved this problem in 1 948, when he invented the field of information theory:

  Each channel has a **capacity**, which is the maximum bit-rate transmission per use of the channel with arbitrarily small probability of error.





Information theory: First homework

- $\boxed{\text{Example 1}}$ **(The weighing problem. (MacKay 4.1))** You are given 12 balls, all equal in weight except for one that is either heavier or lighter.

  You are also given a two-pan balance to use.

  In each use of the balance you may put any number of the 12 balls on the left pan, and the same number on the right pan, and push a button to initiate the weighing; there are three possible outcomes:

  1. either the weights are equal; or
  2. the balls on the left are heavier; or
  3. the balls on the left are lighter.

  Your task is to design a strategy to determine which is the odd ball and whether it is heavier or lighter than the others in as few uses of the balance as possible.

Information theory: First homework

- $\boxed{\text{Example 1}}$ (Continued)

While thinking about this problem, you may find it helpful to consider the following questions:

a) How can we measure information?

b) When you have identified the odd ball and whether it is heavy or light, how much information have you gained?

c) Once you have designed a strategy, draw a tree showing, for each of the possible outcomes of a weighing, what weighing you perform next. At each node in the tree, how much information have the outcomes so far given you, and how much information remains to be gained?

d) How much information is gained when you learn (i) the state of a flipped coin; (ii) the states of two flipped coins; (iii) the outcome when a four-sided die is rolled?

e) How much information is gained on the first step of the weighing problem if 6 balls are weighed against the other 6? How much is gained if 4 are weighed against 4 on the first step, leaving out 4 balls?

Information theory: First homework

- $\boxed{\text{Example 1}}$ (Continued)

  Save your answers to this homework: we will revisit it later in this course.

  First, however, we will use the next few classes to review **Discrete Probability Theory**.