Slides

Introduction

Organization

Professors:

- Lectures: Nicolae Cleju (nikcleju@etti.tuiasi.ro)
- Laboratories: Daniel Matasaru (..@etti.tuiasi.ro)

Grades

$$\text{Final grade} = 0.75 \text{ Exam} + 0.25 \text{ Lab}$$

Time schedule

- 14 weeks of lectures (3h each)
- 14 weeks of laboratories (2h each)
  - 5 laboratories
  - 7 seminars
  - 1 recuperari
  - 1 test

Course structure

1. Introduction to probabilities
2. Pam

Bibliography

1. Pam Pam
2. HamHam
3. Yoyo

Introduction to probabilities

Basic notions of probability

Random variable $=$ the outcome of an experiment

Distribution (probability mass function)

Discrete distribution

Alphabet

Basic properties

Two independent events:

$$P(A \cap B) = P(A) \cdot P(B)$$

Chapter I: Discrete information sources

Block diagram of a communication system

de pus poza

What is information?

Example:

I tell you the following sentence: "your favorite football team lost the last match".

Does this message carry information? How, why, how much?

Consider the following facts:

- the message carries information only because you didn't already know the result.
- if you already known the result, the message is useless (brings no information)
- since you didn't know the result, there were multiple results possible (win, equal or lose)
- the actual information in the message is that *lost* happened, and not *win* or *equal*
- if the result was to be expected, there is little information. If the result is highly unusual, there is more information in this message

Information source

We will always consider information in a context similar to the above example.
We will use terminology from probability theory to define information:

- there is a *probabilistic source* that can produce a number of different *events*.
- each event has a certain probability. We know all the probabilities beforehand.
- at one time, an event is randomly selected according to its probability.
- afterwards, a new message can be selected, and so on $==>$ a stream of messages is produced.

The source is called an *information source* and the selected event is a *message*.
A message carries the information that **it** happened, and not the other possible message events that could have been selected.
The quantity of information is dependent in its probability.

Discrete memoryless source

A discrete memoryless source (DMS) is an information source where the messages are **independent** , i.e. the choice of a message at one time does not depend on what were the previous messages

Each message has a fixed probability. The set of probabilities is the *distribution* of the source.

$$S : \begin{pmatrix} s_1 & s_2 & s_3 \\ \frac{1}{2} & \frac{1}{4} & \frac{1}{4} \end{pmatrix}$$

Properties:

- Discrete: it can take a value from a discrete set (alphabet)
- Complete: $\sum p(s_i) = 1$
- Memoryless: succesive values are independent of previous values (e.g. successive throws of a coin)

A message from a DMS is also called a *random variable* in probabilistics.

Examples

A coin is a discrete memoryless source (DMS) with two messages:

$$S : \begin{pmatrix} coin\_head & coin\_tail \\ \frac{1}{2} & \frac{1}{2} \end{pmatrix}$$

A dice is a discrete memoryless source (DMS) with six messages:

$$S : \begin{pmatrix} s_1 & s_2 & s_3 & s_4 & s_5 & s_6 \\ \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} \end{pmatrix}$$

Playing the lottery can be modeled as DMS:

$$S : \begin{pmatrix} s_1 & s_2 \\ 0.9999 & 0.0001 \end{pmatrix}$$

Examples

An extreme type of DMS containing the certain event:

$$S : \begin{pmatrix} s_1 & s_2 \\ 1 & 0 \end{pmatrix}$$

Receiving an unknown *bit* (0 or 1) with equal probabilities:

$$S : \begin{pmatrix} 0 & 1 \\ \frac{1}{2} & \frac{1}{2} \end{pmatrix}$$

Information

When a DMS provides a new message, it gives out some new information, i.e. the information that a particular message took place.

The information attached to a particular event (message) is rigorously defined as:

$$i(s_i) = -log_2(p(s_i))$$

Properties:

- $i(s_i) \geq 0$
- lower probability (rare events) means higher information
- higher probability (frequent events) means lower information
- a certain event brings no information: $-log(1) = 0$
- an event with probability 0 brings infinite information (but it never happens..)

Entropy of a DMS

We usually don't care about a single message. We are interested in a large number of them (think millions of bits of data).

We are interested in the *average* information of a message from a DMS.

Definition: the entropy of a DMS source S is **the average information of a message**:

$$H(S) = \sum_k p_k i(s_k) = -\sum_k p_k log_2(p_k)$$

where $p_k = p(s_k)$ is the probability of message k.

The choice of logarithm

Any base of logarithm can be used in the definition.

Usual convention: use binary logarithm $log_2()$. H(S) measured in *bits* (*bits / message*)

If using natural logarithm $ln()$, H(S) is measured in *nats*.

Logarithm bases can be converted to/from one another:

$$log_b(x) = \frac{log_a(x)}{log_a(b)}$$

Entropies using different logarithms differ only in scaling:

$$H_b(S) = \frac{H_a(S)}{log_a(b)}$$

Examples

- Coin: $H(S) = 1 bit/message$
- Dice: $H(S) = log(6) bits/message$
- Lottery: $H(S) = -0.9999 log(0.9999) - 0.0001 log(0.0001)$
- Receiving 1 bit: $H(S) = 1 bit/message$ (hence the name!)

Interpretation of the entropy

All the following interpretations of entropy are true:

- H(S) is the *average uncertainty* of the source S
- H(S) is the *average information* of messages from source S
- A long sequence of N messages from S has total information $\approx N \cdot H(S)$
- H(S) is the minimum number of bits (0,1) required to uniquely represent an average message from source S

Properties of entropy

We prove the following properties:

1. $H(S) \geq 0$ (non-negative)

2. $H(S)$ is maximum when all n messages have equal probability $\frac{1}{n}$. The maximum value is $maxH(S) = log(n)$.

3. *Diversfication* of the source always increases the entropy

Example - Game

Game: I think of a number between 1 and 8. You have to guess it by asking yes/no questions.

- How much uncertainty does the problem have?
- How is the best way to ask questions? Why?
- What if the questions are not asked in the best way?
- On average, what is the number of questions required to find the number?

Example - Game v2

Suppose I choose a number according to the following distribution:

$$S : \begin{pmatrix} s_1 & s_2 & s_3 & s_4 \\ \frac{1}{2} & \frac{1}{4} & \frac{1}{8} & \frac{1}{8} \end{pmatrix}$$

- On average, what is the number of questions required to find the number?
- What questions would you ask?

What if the distribution is:

$$S : \begin{pmatrix} s_1 & s_2 & s_3 & s_4 \\ 0.14 & 0.29 & 0.4 & 0.17 \end{pmatrix}$$

In general:

- What distribution makes guessing the number the most difficult?
- What distribution makes guessing the number the easiest?

Information flow of a DMS

Suppose that message s_i takes time t_i to be transmitted via some channel.

Definition: the information flow of a DMS S is **the average information transmitted per unit of time**:

$$H_\tau(S) = \frac{H(S)}{\overline{t}}$$

where $\overline{t}$ is the average duration of transmitting a message:

$$H(S) = \sum_i p_i t_i$$

Extended DMS

Definition: the n-th order extension of a DMS S, S^n is the source with messages has as messages all the combinations of n messages of S:

$$\sigma_i = \underbrace{s_j s_k \ldots s_l}_{n}$$

- If S has k messages, S^n has k^n messages
- Since S is DMS

$$p(\sigma_i) = p(s_j) \cdot p(s_k) \cdot \ldots \cdot p(s_l)$$

Extended DMS - Example

Examples:

$$S : \begin{pmatrix} s_1 & s_2 \\ \frac{1}{4} & \frac{3}{4} \end{pmatrix}$$

$$S^2 : \begin{pmatrix} \sigma_1 = s_1 s_1 & \sigma_2 = s_1 s_2 & \sigma_3 = s_2 s_1 & \sigma_4 = s_2 s_2 \\ \frac{1}{16} & \frac{3}{16} & \frac{3}{16} & \frac{9}{16} \end{pmatrix}$$

$$S^3 : \begin{pmatrix} s_1 s_1 s_1 & s_1 s_1 s_2 & s_1 s_2 s_1 & s_1 s_2 s_2 & s_2 s_1 s_1 & s_2 s_1 s_2 & s_2 s_2 s_1 & s_2 s_2 s_2 \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots \end{pmatrix}$$

Extended DMS - Another example

Long sequence of binary messages:

010011001110010100...

Can be grouped in bits, half-bytes, bytes, 16-bit words, 32-bit long words, and so on.

Property of DMS

We prove the following:

Theorem: The entropy of a n-th order extension is n times larger than the entropy of the original DMS

$$H(S^n) = nH(S)$$

Interpretation: grouping messages from a long sequence in blocks of n does not change total information (e.g. groups of 8 bits = 1 byte)

Sources with memory

An example

The distribution (frequencies) of letters in English:

letter	probability	letter	probability
A	.082	N	.067
B	.015	O	.075
C	.028	P	.019
D	.043	Q	.001
E	.127	R	.060
F	.022	S	.063
G	.020	T	.091
H	.061	U	.028
I	.070	V	.010
J	.002	W	.023
K	.008	X	.001
L	.040	Y	.020
M	.024	Z	.001

Text from a memoryless source with these probabilities:

XFOML RXKHRJFFJUJ ZLPWCFWKCYJ
FFJEYVKCQSGXYD QPAAMKBZAACIBZLHJQD

(taken from Elements of Information Theory, Cover, Thomas)

What's wrong? **Memoryless**

Sources with memory

Definition: A source has memory of order m if the probability of a message depends on the last m messages.

The last m messages = the **state** of the source (S_i).

A source with n messages and memory $m \Rightarrow n^m$ states in all.

For every state, messages can have a different set of probabilities.

Also known as Markov sources.

Transitions

When a new message is provided, the source **transitions** to a new state:

$$\underbrace{s_i s_j s_k}_{\text{old state}} \; s_l$$

$$s_i \; \underbrace{s_j s_k s_l}_{\text{new state}}$$

The message probabilities = the probabilities of transitions from state S_a to S_b

Transition matrix

The transition probabilities are organized in a **transition matrix** $[T]$

$$[T] = \begin{bmatrix} p_{11} & p_{12} & \dots & p_{1N} \\ p_{21} & p_{22} & \dots & p_{2N} \\ \dots & \dots & \dots & \dots \\ p_{N1} & p_{N2} & \dots & p_{NN} \end{bmatrix}$$

p_{ij} is the transition probability from state S_i to state S_j

N is the total number of states

Graphical representation

Example here