Int. J. Engng Ed. Vol. 26, No. 3, pp. 602–611, 2010
Printed in Great Britain.
0949-149X/91 $3.00+0.00

An Educational Methodology to Interpret the Entropy of a Source Based on the Analogy between such an Entropy and the Mechanical Energy*

JORGE BONACHE, CARMEN COUSIDO and WILMAR HERNANDEZ
EUIT de Telecomunicacion, Universidad Politecnica de Madrid, Campus Sur UPM, 28031 Madrid, Spain.
E-mails: jbonache@euitt.upm.es; ccousido@euitt.upm.es; whernan@ics.upm.es

This paper describes a seminar within the work unit 'Introduction to the Entropy in Communications' in the Statistics and Stochastic Processes course. During the seminar the students study the entropy of a discrete source, its interpretation as potential information, and the parallelism with the law of conservation of energy. A detailed analysis of the singularity that represents the case of a uniform source distribution is carried out and analogies made with the principle of conservation of mechanical energy. In addition, the intuitive characteristics of the seminar are highlighted, as well as its gradually increasing development, carrying out successive generalizations. The methodologies used in the seminar allow the students to discover for themselves an interdisciplinary model of learning as well as showing them how to use computer simulation as a learning strategy in order to achieve the learning objectives.

Keywords: entropy; information theory; law of conservation of energy; computer simulation

1. INTRODUCTION

1.1 Entropy in communication

TWO OF THE MOST IMPORTANT ISSUES in communication theory are (1) determining what the ultimate data compression is and (2) determining the ultimate transmission rate of communication, in accordance with [1]. The information theory based on the work of Shannon [2, 3] made significant contributions to the solution to the above issues. The theory developed by Shannon uses probability distributions to quantify the information through the entropy function. With regard to the first issue, Shannon relates the entropy to the average length of the words used to code the information. The entropy represents the infimum of such a length.

With regard to the second issue, the transmission of information through a noisy channel distorts the information and introduces errors. Conditional probabilities allows one to define a set of entropies that measure the information from the transmitter and receiver points of view, and the capacity of the channel is defined from those entropies.

Shannon [2, 3] establishes that, regardless of the presence of noise, the capacity of a channel is the supremum of the rates at which the information can be transmitted with an arbitrarily small probability of error.

According to [5], the concept of entropy proposed by Shannon (1948) was based on the entropy of Boltzmann (1896), from statistical mechanics. Boltzmann was the first to state the probabilistic meaning of the entropy in thermodynamics, because of that he is considered to be the precursor of information theory. He interpreted the entropy of a physical system as a measure of its disorder. In a physical system with many degrees of freedom the number that measures the disorder of the system also measures the uncertainty of the individual states of the particles.

Some decades before Clausius (1864) had defined the entropy of a system in terms of its measurable characteristics, Boltzmann realized its probabilistic meaning.

In analogy with the expression of Boltzmann, in 1948 Shannon introduced the entropy in abstract form as a measure of the amount of information or uncertainty that a random experiment contributes, through the concept of probability.

Hartley (1928) tried to define a measure of the amount of information but such a measure only took into account the number of results of the experiment, instead of its probabilities. The achievement of Shannon was to extend the concept so that the entropy of Hartley was a particular case of that of Shannon when the results of the experiment have the same probability.

Because the entropy is a basic piece of the information theory and because it intervenes directly in the calculation of the limits that exist in communication with regard to data compression and data transmission, we consider it to be of

* Accepted 19 December 2009.

fundamental importance for students of telecommunication to carry out a deep study of the concept of entropy and its interpretation.

The reality is that, on the one hand, in both science and engineering we find a lot of applications of the concept of entropy aimed at solving specific real-life problems [6–11].

However, on the other hand, in engineering education, the teaching of entropy to undergraduate students in engineering schools focuses basically on using it in some specific applications.

1.2 Seminar. Objectives and methodologies

Taking the above statements into consideration, from the point of view of Information Theory Education and with the aim of strengthening the theoretical basis of undergraduate telecommunication students on communication theory, in this paper we present a seminar on both the interpretation of the entropy of a source as a central concept in information theory and communications, and its analogies with mechanical energy.

At this point, it should be highlighted that the analysis of such analogies is a novel approach that allows telecommunication students to learn about more examples of the close relationship that exists among topics of subjects that are apparently very distant from each other, such as 'Mechanics' and 'Statistics and Stochastic Processes'.

The above-mentioned seminar is taught within the subject 'Statistics and Stochastic Processes', at the end of the first semester of the second academic year of the undergraduate students of telecommunication engineering.

The seminar is taught once the students have acquired and applied knowledge of the following:

- Probability. Conditional probability
- Discrete one-dimensional random variables. Expected value
- Computer simulation of random phenomena
- Estimation theory.

The main contributions of this paper are:

- to introduce concepts and results in an intuitive, gradually increasing manner;
- to relate quantitatively lost uncertainty and earned information in a source;
- to identify the couple entropy/information with the couple potential energy/kinetic energy in order to analyze the transformation of uncertainty into information, in a similar manner to the conversion of potential energy into kinetic energy.
- With this philosophy and when the distribution of the source is uniform, we establish a principle that is equivalent to that of conversion of mechanical energy when working in conservative force fields. The generalization of the distribution leads us to a general law of conservation of energy;
- to use computer simulation as a tool that allows students to carry out some experiments and to promote a better intuitive understanding.

This seminar is intended to be taught in 5 hours and is aimed at second-year undergraduate telecommunication students.

First, we expect that it will serve the students as an introduction to information theory.

Second, we expect to elaborate methodologies that allow us to show the students a way of facing and solving new problems by using scientific approximation.

The learning objectives are:

- to highlight the characteristic of potential information of the entropy of a random discrete source;
- to use the principle of conservation of mechanical energy in order to analyze the duality uncertainty/information of the entropy of a source;
- to establish some basis that allows the students to better understand successive generalizations of the entropy (conditional entropy, entropy of two or more random variables) as well as concepts that come from them (mutual information and channel capacity).

The methodologies are:

M1. Approximation of a concept through a particular case of study
M2. Use of the intuition to make a first interpretation of the entropy
M3. Consecutive generalizations of the initial conditions.
M4. Search for models in other fields of engineering and science
M5. Use of computer simulation in order to corroborate results, search for counterexamples and establish conjectures
M6. Proposal of a series of key questions whose answers allow one to advance in the study and in reaching conclusions.

In Section 2.1 we laid out the basic knowledge about entropy that the students will learn in order to understand the seminar. Also, taking an interpretation of the entropy of the source as the uncertainty of the symbol to be transmitted, we propose to gradually incorporate information about the transmitted symbol and to analyze the value of the addition of the amount of acquired information and uncertainty that remains.

In Section 2.2 an introductory example is given, and Section 2.3 is devoted to a study of the case in which the probability distribution of the source is uniform. In Section 2.3, we highlight that the above-mentioned addition is constant when the information that we have about the emitted symbol varies.

In Section 2.4, we interpret this case of study as a conversion of uncertainty into amount of information similar to the one that is produced between potential energy and kinetic energy in conservative force fields. We establish a mathematical expression that is equivalent to the principle of conservation of mechanical energy in fields that come from a potential.

In Section 2.5, we use computer simulation to give an empirical proof of the above result and in Section 2.6 we see that it is impossible to generalize this result to an arbitrary probability distribution.

In Section 2.7, we search for a mathematical expression for the case of a source with an arbitrary probability distribution, and in Section 2.8 we carry out its interpretation and establish analogies with the law of conservation of energy in non-conservative fields. In Section 2.9 we carry out an assessment of the seminar. The last section, Section 3, is devoted to our conclusions.

2. DESCRIPTION OF THE SEMINAR

2.1 Basic concepts

We begin the seminar by explaining the definition of information contained in an event [12].

Definition 1: Let A be an event of a probabilistic space (Ω, S, P), the amount of information that is gained by knowing that the event A has occurred is given by

$$I(A) = \log_a \frac{1}{P(A)} = -\log_a P(A), P(A) > 0$$

Next, we define a measure of the uncertainty associated to a discrete random variable [1].

Definition 2 (Shannon entropy): If X is a discrete random variable that takes the values x_1, x_2, ..., x_n with probabilities $p_i = P(X = x_i)$, $i = 1, 2, \ldots. n$, the entropy of the random variable X is given by

$$H(X) = -\sum_{i=1}^{n} p_i \log_a p_i \tag{1}$$

The indeterminate form $p_i \cdot \log_a p_i$ with $p_i = 0$ is solved by defining

$$p_i \cdot \log_a\ p_i = 0 \textit{ if } p_i = 0$$

Also, as $I(X = x_i) = \log_a \frac{1}{p_i}$, we can write

$$H(X) = \sum_{i=1}^{n} p_i I(X = x_i) \tag{2}$$

That is, $H(X)$ is the average information contributed by the events $(X = x_i)$, $i = 1, \ldots n$.

If $a = 2$, the entropy is expressed in *bits*, which is the entropy of a random variable that takes only two values with identical probability.

From now on we will consider $a = 2$, because we are in the age of digital communications and almost all the information is digitally transmitted.

In accordance with [13], the following proposition shows the properties of the entropy of a discrete random variable.

Proposition 1: Let X be a discrete random variable that takes the values x_1, x_2, ..., x_n with probabilities $p_i = P(X = x_i)$, $i = 1, 2, \ldots n$. We then find that

1) $H(X)$ is continuous in the probabilities p_i.
2) $H(X)$ is symmetric:
$$H(p_1, ..., p_n) = H\left(p_{\sigma(1)}, ..., p_{\sigma(n)}\right),$$
where σ is any permutation of the subscripts (1, ..., n).
3) $0 \leq H(X) \leq \log_2 n$
   a. $H(X) = 0$ if and only if $p_k = 1$ for some k (so that $p_i = 0$ for all $i \neq k$).
   b. $H(X) = \log_2 n$ if and only if $p_i = 1/n$ for $i = 1, ..., n$.
4) $H(X)$ increases as n increases if X is a uniform random variable.

At this point, the students are asked to do the exercise below as homework assignment (30 minutes, individual work after class).

EXERCISE 1: Prove *Proposition* 1.

Finally, we define the meaning of both source and entropy of a source [14].

Definition 3: A source is an ordered pair (S, P) where $S = \{x_1, ..., x_n\}$ is called the source alphabet and P is a probability distribution in S. The probability of x_i is denoted with p_i, $i = 1, ..., n$.

Definition 4 (Entropy of a source): Let (S, P) be a source. The entropy of a source is defined as the entropy of the random variable X that takes the values x_1, x_2, ... , x_n with probabilities p_i, $i = 1, 2, ..., n$, that is

$$H(X) = -\sum_{i=1}^{n} p_i \log_2 p_i$$

At this point, the introductory example shown below is given.

2.2 Introductory example: A source that emits three symbols with uniform probability distribution—Methodologies M1 and M2

We consider a source that emits one of the three symbols 1, 2 or 3, with probabilities $p_i = 1/3$, $i = 1, 2, 3, H \equiv H(X) = \log_2 3$.

We have a total and maximum uncertainty with respect to the symbol that is going to be emitted, because the three symbols have the same probability of being emitted.

Let us assume that we receive the information that the emitted symbol is odd. Let B be the vent, the emitted digit is odd.

The occurrence of B contributes the following amount of information: $I(B) = -\log_2 \frac{2}{3}$.

The information contributed by the occurrence of B has decreased the uncertainty with respect to what have been emitted (entropy of the source). Nevertheless, such information has not made the uncertainty zero because now there are two equiprobable results, 1 and 3.

How much does the entropy have decreased?

When we incorporate the information, the probabilities of 1, 2 and 3 are 1/2, 0 and 1/2, respectively. Thus, the entropy of the source will be $H_B \equiv H(X/B) = \log_2 2$ and the loss of uncertainty will be $H - H_B = -\log_2 \frac{2}{3}$.

We observe that the loss of uncertainty coincides with the information contained in B. Thus this has been useful for decreasing the uncertainty (entropy) of the source.

Let us assume now that the information we receive with respect to the emitted symbol is that the emitted symbol is even. Let A be the event, the emitted symbol is even. The information contained in A is $I(A) = \log_2 3$.

With the information contributed by A the uncertainty of the source is zero because the only possibility is that the emitted symbol is 2. This event totally specifies the result.

Let H_A be the entropy that the source has if we know that A has occurred. As now only the symbol 2 has a probability different from 0 and this is equal to 1, we have that $H_A = 0$. The loss of uncertainty will be $H - H_A = \log_2 3 - 0 = I(A)$. Then, we conclude that

$$H = H_B + I(B) = 0 + I(A) = I(A)$$

In the two cases that have been analyzed so far, the entropy that the source has, after having had both information on the event that determines partially or totally the result and information on the amount of information contained in such an event, is a constant and is equal to the entropy of the source.

2.3 Generalization to a source with arbitrary uniform distribution—Methodology M3

At this point, it is important to verify whether the result obtained in Section 2.2 is true for an arbitrary uniform distribution and for any event.

We will consider a source that emits one of the symbols x_1, ..., x_n with probability $p_i = 1/n$, $i = 1, 2, ..., n$.

As we have been doing up to now, we will continue using the following notation:

- H entropy of the source
- B event that provides partial information on the emitted symbol
- H_B entropy of the source after having had information on the occurrence of B
- A event that determines what the emitted symbol was.
- H_A entropy of the source after having known what the emitted symbol was.

The entropy of the source will be:

$$H = -\sum_{i=1}^{n} \frac{1}{n} \log_2 \frac{1}{n} = \log_2 n \qquad (3)$$

Assume that we know that the event B has occurred; it has emitted one of the m symbols $x_{i1}, ..., x_{im}$, for $m < n$. Let p'_i be the probability of emitting the symbol i given that B has occurred, that is,

$$p'_i = P(X = x_i/B) = \frac{1}{m}$$

The entropy of the source having known that B has occurred is

$$\begin{aligned} H_B &= -\sum_{i=1}^{m} p'_i \log_2 p'_i \\ &= -\sum_{i=1}^{m} \frac{1}{m} \log_2 \frac{1}{m} = \log_2 m \qquad (4) \end{aligned}$$

and the information contributed by the event B is

$$I(B) = -\log_2 \frac{m}{n} = \log_2 \frac{n}{m} \qquad (5)$$

Therefore, we verify that

$$H_B + I(B) = \log_2 m + \log_2 \frac{n}{m} = \log_2 n \qquad (6)$$

Thus, $H_B + I(B)$ depends only on the number of symbols that the source can emit.

If we know what the emitted symbol was and call this event A, then the entropy H_A will be equal to zero and

$$I(A) = -\log_2 \frac{1}{n} = \log_2 n \qquad (7)$$

Consequently we can establish that

$$H = H_B + I(B) = H_A + I(A) = I(A) \qquad (8)$$

$H = I(A)$ indicates that the entropy of the source (uncertainty with respect to the emitted symbol) has been transformed into information (the information acquired when we know what the emitted symbol was).

Furthermore, $H = H_B + I(B)$, that is, the loss of uncertainty is converted into earned information. So that in that process the addition of the entropy that the source has altered, having known an event that determines partially or totally the result and the amount of information that such an event contains, is a constant.

2.4 Interpretation in terms of energy—Methodology M4

Let us now look for a model with which we can identify the previous result. To this end, we consider the framework of Mechanics in Physics: in that framework a conservation of equivalent magnitudes is produced [15].

A particle in motion, under only the action of conservative forces, $\vec{F} = \nabla U$, has a potential energy U owing to its position, and a kinetic energy W owing to its speed.

During the movement there is a conversion from one type of energy to the other, in such a way that the mechanical energy, which is the addition of both energies, is kept constant.

Let us assume that such a particle, placed at some distance (P) from the floor and under the action of the gravitational force ($\vec{F}$), falls without the effect of the resistance due to the air. Its potential energy decreases and its kinetic energy increases in the same manner (Q), in such a way that when the particle reaches the floor (R) all the potential energy has been transformed into kinetic energy.

In accordance with the law of conservation of mechanical energy, when the particle is under only the action of conservative forces such as the gravitational force, the mechanical energy remains constant along the trajectory.

Let us establish the analogy between the conversion from one energy into another, and what we explained in previous subsections about the entropy of a source.

The entropy of the source plays the role of the potential energy. This decreases as the partial information on the emitted symbol increases. When we know what the emitted symbol was, the entropy converts itself into information. The amount of information available (i.e., actual information) plays the role of the kinetic energy, in such a way that the addition of both energies is kept constant throughout the process.

Then, we can consider that the entropy has the nature of potential information, due to its capacity to produce actual information.

Equation (8) can be interpreted in the following way: The sum of the potential and actual information does not change when the distribution of the source is uniform. Fig. 1 shows this analogy.

Therefore, Equation (8) is equivalent to the principle of conservation of mechanical energy.

2.5 Empirical proof of the previous result—Methodology M5

Let us assume that a source can emit the symbols $i = 1, 2, \ldots, 10$, with $p_i = 1/10$. Then, substituting $n = 0$ and $m = 5$ in (3), (4), (5) and (7), we obtain the results shown in Table 1 for the following cases:

i) When we do not have any information on the emitted symbol.
ii) When we know that the emitted symbol is even (event B).
iii) When we know what the emitted symbol was (event A).

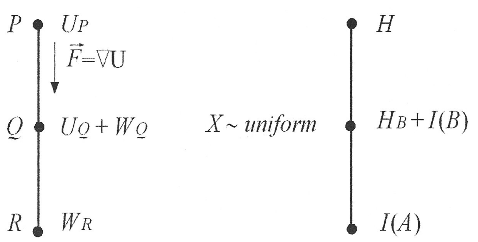

Fig. 1. Law of conservation of mechanical energy versus entropy.

Let us now generate a sample of a discrete uniform variable X in $\{1, \ldots, 10\}$ and estimate the entropy and information from such a sample. Let us use the relative frequency of each symbol f_{r_i} as an estimator of the probability p_i [16]. Then we estimate the entropy of the random variable X by using the statistic

$$\hat{H} = -\sum_{i=1}^{10} f_{r_i} \log_2(f_{r_i}) \tag{9}$$

where f_{r_i} is the relative frequency at each digit i, for $i = 1, \ldots, 10$, of the sample. The information contained in each symbol i, for $i = 1, \ldots, 10$, is estimated by using the statistic

$$\hat{I}(A_i) = -\log_2(f_{r_i}) \tag{10}$$

The information contained in B is estimated by using the statistic

$$\hat{I}(B) = -\log_2(f_r(B)) \tag{11}$$

where $f_r(B)$ is the relative frequency of B in the sample.

The entropy that the source has, after the occurrence of B, is estimated by

$$\hat{H}_B = -\sum_{i=1}^{5} \frac{f_{r2i}}{f_r(B)} \log_2 \frac{f_{r2i}}{f_r(B)} \tag{12}$$

where $f_{r_{2i}}$ is the relative frequency even digit of the sample.

Then, we carry out some computer simulations by using R [17, 18], because R has the advantage of being freely redistributable and widely used in Statistics.

However, as mentioned in [19] we can use other platforms such as Matlab that are more general purpose numerical engines than R. But such platforms tend to have less support for certain statistical functions.

The above-mentioned simulations need only a few lines of code. The function entropy of Appendix A carries out the calculation of the statistics $\hat{H}$, $\hat{I}(B)$, $\hat{H}_B$, $Suma = \hat{H}_B + \hat{I}(B)$ and $error = |Suma - \hat{H}|$ in accordance with (9), (11) and (12), from a single random sample of a distribution of probability $p = (p_1, \ldots p_{10})$. We save the **function** entropy in the text file entropy.R. In order to carry out the simulation process, we execute the following R code:

Table 1. Theoretical entropy

H	H_B	*I(B)*	*I(A)*
$\log_2 10$	$\log_2 5$	$\log_2 \frac{10}{5}$	$\log_2 10$

```
>source("entropy.R")
>entropy()
>entropy(1000)
>entropy(10000)
```

Table 2 shows the results obtained for $s = 100$, $s = 1000$ and $s = 10\,000$.

At this point, we divide the seminar into several small groups consisting of three students (each group) and we ask each group to solve the exercise below as a homework assignment (after class in no more than 15 minutes).

EXERCISE 2: Let $n = 5$, $p = (0.2, 0.2, 0.2, 0.2, 0.2)$ and the event B, the emitted digit is odd. Fill in Table 2 for this case.

2.6 Attempt to generalize (8) to the case of an arbitrary distribution—Methodologies M3 and M5

Will $H_B + I(B)$ be a constant if we consider a non-uniform source distribution?

Let us consider a simple random sample of one discrete random variable X that takes the values in $\{1, ..., 9\}$ with probabilities $p_1 = 0.1$, $p_2 = 0.3$, $p_3 = 0.05$, $p_4 = 0.2$, $p_5 = 0.15$, $P_6 = 0.12$, $p_7 = 0.04$, $p_8 = 0.02$ and $p_9 = 0.02$.

Table 3 shows the results obtained in the simulations for s = 100, s = 1000 and s = 10 000, choosing $p = (p_1, p_2, p_3, p_4, p_5, p_6, p_7, p_8, p_9)$.

We execute the following R code:

```
>p = c(0.1, 0.3 , 0.05, 0.2, 0.15, 0.12, 0.04, 0.02, 0.02)
>sum(p)==1
>entropy(100,9,p)
>entropy(1000,9,p)
>entropy(10000,9,p)
```

In this case, we observe that

$$\begin{aligned} H &= -\sum_{i=1}^{9} p_i \log_2 p_i = 2.722 \\ H_B &= -\sum_{i=1}^{4} \frac{p_{2i}}{P(B)} \log_2 \frac{p_{2i}}{P(B)} = 1.645 \\ I(B) &= -\log_2 P(B) = 0.643. \end{aligned} \tag{13}$$

At this point, we divide the seminar into small groups of three students one more time and ask them to solve the exercise below as a homework assignment. Also, they have to carry out a comparative analysis among their results. The estimated time to do this exercise is about 25 minutes.

EXERCISE 3: In this exercise both n and the event B are chosen arbitrarily. Then, fill in Table 3 for this case.

2.7 Searching for an expression of HB + I(B) for the case of an arbitrary distribution source—Methodologies M3, M5 and M6

From the previous empirical results we establish the following conjecture [20]:

Conjecture: The sum of the potential and actual information is not a constant if the distribution of the source is not uniform.

Is there any reason that can assure us that the above conjecture is true?

For the equiprobable case, we said that the entropy of the source has been transformed into the information that contains the knowledge that we have on the symbol that has been emitted. That is, $H = I(X = x_i)$.

When the probabilities of the symbols are not equal to each other, the information of the elemental events $(X = x_i)$ are not equal either and the equality cannot be true.

Let us try to search for an expression for the sum of the potential and actual information, $H_B + I(B)$.

In order to do this, we consider a source that emits one of the n symbols x_1, ..., x_n, with probabilities p_1, ..., p_n, respectively, and

$$\sum_{i=1}^{n} p_i = 1.$$

Let us consider the events A_i, the symbol x_i, for $i = 1, ..., n$, is emitted.

The information contributed by the occurrence of A_i is, $I(A_i) = -\log_2 p_i \;\; i = 1, ..., n$.

The entropy of the source will be

$$H = \sum_{i=1}^{n} p_i I(A_i) \tag{14}$$

We know that it can only be emitted one of the m symbols (event B) x_{i1} ...... x_{im}, with probabilities p_{i1} ...... p_{im}.

The occurrence of B provides the amount of information

$$I(B) = -\log_2(p_{i1} + ... + p_{im}) \tag{15}$$

and the entropy that the source has after having known B is

$$H_B = -\sum_{k=1}^{m} p'_{ik} \log_2 p'_{ik} \tag{16}$$

Table 2. Estimated entropy: uniform case

	$\hat{H}$	$\hat{H}_B$	$\hat{I}(B)$	***Suma***	***Error***
s = 100	3.279	2.294	0.785	3.080	0.149
s = 1000	3.314	2.318	0.971	3.290	0.025
s = 10000	3.321	2.321	1.004	3.325	0.004

Table 3. Estimated entropy: non-uniform case

	$\hat{H}$	$\hat{H}_B$	$\hat{I}(B)$	***Suma***	***Error***
s = 100	2.595	1.347	0.736	2.084	0.511
s = 1000	2.740	1.665	0.662	2.327	0.412
s = 10000	2.725	1.647	0.643	2.290	0.434

where p'_{i_k} is the probability of emitting the symbol i_k given the occurrence of B, that is

$$p'_{ik} = P(X = x_{ik}/B) = \frac{P(X = x_{ik})}{P(B)} = \frac{p_{ik}}{P(B)} \quad (17)$$

Then,

$$\begin{aligned} H_B &= -\sum_{k=1}^{m} \frac{p_{ik}}{P(B)} \log_2 \frac{p_{ik}}{P(B)} \\ &= -\frac{1}{P(B)} \sum_{k=1}^{m} p_{ik} (\log_2 p_{ik} - \log_2 P(B)) \\ &= -\frac{1}{P(B)} \left(\sum_{k=1}^{m} p_{ik} \log_2 p_{ik} - \log_2 P(B) \sum_{k=1}^{m} p_{ik} \right) \\ &= -\frac{1}{P(B)} \left(\sum_{k=1}^{m} p_{ik} \log_2 p_{ik} - P(B) \log_2 P(B) \right) \\ &= \log_2 P(B) - \sum_{k=1}^{m} \frac{p_{ik}}{P(B)} \log_2 p_{ik} \\ &= -I(B) + \sum_{k=1}^{m} p'_{ik} I(A_{ik}) \end{aligned} \quad (18)$$

Therefore,

$$H_B + I(B) = \sum_{k=1}^{m} p'_{ik} I(A_{ik}) = \sum_{i=1}^{n} P(X = x_i/B) I(A_i) \quad (19)$$

where $P(X = x_i/B) = 0$, for $i \neq i_1, i_2, \ldots, i_m$.

Equation (19) tells us that the sum of both the entropy that the source has after the occurrence of B and the information that B contributes is a weighted average of the information contained in the symbols to be emitted x_1 ...... x_n, where the weights are the respective probabilities of each one of the symbols after the occurrence of B. Therefore, (19) is an expression for the sum of potential and actual information, and represents an extension of (8) for the case of an arbitrary distribution of the source.

At this point, we ask the students to complete the exercise below individually as a homework assignment. The estimated time to do this exercise is about 20 minutes.

EXERCISE 4: Show that for the particular case of uniform distribution, for the symbols x_1 ..... x_n, the result obtained in (19) coincides with the one that we obtained at the beginning of our study, in (8).

2.8 Interpretation in terms of energy—Methodology M4

By identifying entropy with potential energy and actual information with kinetic energy, as we have done so far, unlike the case discussed in Section 2.3, we have obtained a non-constant expression for the mechanical energy when the source has an arbitrary distribution. However, the mechanical energy changes due to the existence of non-conservative forces, and the law of conservation of energy states that the increase/decrease in the amount of mechanical energy of a system is equal to the decrease/increase in the amount of its internal energy.

For the case in which there are only conservative forces, there is no variation in the internal energy: this law is the principle of conservation of mechanical energy.

In the absence of information on the symbol emitted by the source, the mechanical energy will be the entropy H. However, when we know what the emitted symbol x_i is, there is no entropy left and the mechanical energy will be the amount of information of the symbol, $I(A_i)$.

$$H - I(A_i) = \sum_{j=1}^{n} p_j I(A_j) - I(A_i) \quad (24)$$

Then, if $I(A_j) < H$, the amount of information of $A_i = (X = x_i)$ is less than the amount of the average information of the symbols. Not all the entropy of the source has been converted into the amount of information of the event A_i. There is a loss of mechanical energy of the system that is converted into earned internal energy.

If $H < I(A_i)$, the amount of information of A_i is greater than the amount of the average information of the symbols. Mechanical energy has been earned at the expense of a loss of the same amount in internal energy.

If for all values of i, $i = 1 \ldots n$, we have that $H - I(A_i) = 0$, then $p_i = 1/n$ and the probability distribution of the source is uniform, as we already knew.

If for all values of i, $i = 1 \ldots n$, $H - I(A_i)$ takes values close to zero, the amount of information that the events A_i contain are all close to the average and therefore very close to each other.

As a consequence, the probabilities of the symbols are close to each other and therefore close to $1/n$. This means that there is a certain 'similarity' between the distribution of probabilities $p_1, \ldots, p_n$ and the uniform distribution. The loss or earn of internal energy when the different symbols are emitted is small. That is, the probabilities of the source are 'close' to the ones of the uniform distribution with the same number of symbols, and we approach the model of conservation of mechanical energy.

2.9 Assessment

To end the seminar we are going to measure the learning and retention of the material [21–24]. To that end, several key issues are going to be taken into consideration, such as: performance of the students in the seminar, teamwork, and so on.

During the last 30 minutes of the seminar, we are going to conduct an Assessment Test, in which we are going to measure the individual work of the students. In that test we are going to ask the students to answer some important multiple answer questions based on the exercises of the homework assignments 1 and 4, and the interpretation of the results (8) and (19). An example of the full version of one of these Assessment Tests is included in Appendix B.

Furthermore, the teamwork for the homework-assignment Exercises 2 and 3 is also measured.

3. CONCLUSIONS

In this paper a seminar on the interpretation of the entropy of a source and its analogies with energy has been presented. Owing to its characteristics the seminar has been aimed mainly at telecommunication students who are interested in learning about the entropy in communications.

The character of the potential information of the entropy has been highlighted. Also, the duality uncertainty/information of the entropy of a source has been analyzed, and the basis that will allow us to carry out further generalizations of the entropy and define concepts derived from it have also been established.

All the above-mentioned work has been carried out using methodologies that have allowed the student to understand a way of approximating the solution to scientific problems.

Acknowledgements—The authors would like to thank both Professor J. G. Palomo and G. Perez at the Universidad Politecnica de Madrid for encouraging us to write this paper and for their valuable comments and suggestions about it. This research was supported by the Universidad Politecnica de Madrid under the research projects on Engineering Education IE08590220 and IE09590203.

REFERENCES

1. T. M. Cover and J. B. Thomas, *Elements of Information Theory*, 2nd edn, John Wiley & Sons, New Jersey, 2006.
2. C. E. Shannon, A mathematical theory of communication. Bell *System Technical Journal*, **27**, 1949, pp. 379–423 and pp. 623–656.
3. C. E. Shannon and W. Weaver, *The Mathematical Theory of Communication*, University of Illinois Press, Urbana, 1998.
4. G. A. Jones and J. J. Jones, *Information and Coding Theory*, Springer-Verlag, London, 2000.
5. J. van der Lubbe, *Information Theory*, Cambridge University Press, Cambridge, 1997.
6. J. F. Bercher and C. Vignat, Estimating the entropy of a signal with applications, IEEE *Transactions on Signal Processing*, **48**(6), 2000, pp. 1687–1694.
7. O. S. Jahromi, B. A. Francis, and R. W. Kwong, Spectrum estimation using multirate observations, *IEEE Transactions on Signal Processing*, **52**(7), 2004, pp. 1878–1890.
8. E. Learned-Miller and J. DeStefano, A probability upper bound on differential entropy, *IEEE Transactions on Information Theory*, **54**(11), 2008, pp. 5223–5230.
9. J. F. Arnold and M. C. Cavenor, A practical course in digital video communications based on MATLAB, *IEEE Transactions on Education*, **39**(2), 1996, pp. 127–136.
10. R. Kuc, Teaching the nonscience major: EE101–The Digital Information Age, *IEEE Transactions on Education*, **44**(2), 2001, pp. 158–164.
11. S. C. Gustafson, C. S. Costello, E. C. Like, S. J. Pierce and K. N. Shenoy, Bayesian threshold estimation, *IEEE Transactions on Education*, **52**(4), 2009, pp. 400–403.
12. C. W. Therrien and M. Tummmala, *Probability for Electrical and Computer Engineers*, CRC Press LLC, Boca Raton, 2004.
13. L. Pardo, *Teoría de la Información Estadística*, Hespérides, Salamanca, 1998.
14. S. Roman, *Coding and Information Theory*, Springer-Verlag, New York, 1992.
15. H. D. Young and R. A. Fredman, *Sears and Zemansky's University Physics with Modern Physics*, 11th edn, Pearson/Addison-Wesley, San Francisco, 2003.
16. F. M. Deeking C. Kraaikamp H. P. Louhaä and L. E. Meesters, *A Modern Introduction to Probability and Statistics*, Springer-Verlag, London, 2005.
17. *The R Project for Statistical Computing*, www.r-project.org, Accessed 29 October 2009.
18. J. R. Verzani, *Using R for Introductory Statistics*, Chapman & Hall/CRC, Boca Raton, 2005.
19. N. J. Horton, E. R. Brown and L. Qian, Use of R as a toolbox for mathematical statistics exploration, *The American Statistician*, **58**(4), 2004, pp. 343–357.
20. S. Kay, *Intuitive Probability and Random Processes using MATLAB*, Springer, New York, 2006.
21. C. J. Finelli and M. A. Wicks, An instrument for assessing the effectiveness of the circuits curriculum in an electrical engineering program, *IEEE Transactions on Education*, **43**(2), 2000, pp. 137–142.
22. B. Oakley, A, Virtual classroom approach to teaching circuit analysis, *IEEE Transactions on Education*, **39**(3), 1996, pp. 287–296.
23. J. C. Schlimmer, J. B. Fletcher and L. A. Hermens, Team-oriented software practicum, *IEEE Transactions on Education*, **37**(2), 1994, pp. 212–220.
24. C. Connolly, E. Murphy and S. Moore, Programming anxiety amongst computing students—A key in the retention debate, *IEEE Transactions on Education*, **52**(1), 2009, pp. 52–56.

APPENDIX A. FUNCTION ENTROPY

```
entropy = function(s = 100, n = 10, p = rep(1/n,n))
# This function carries out the estimation of
# H, H_B, I(B), Suma and error
# from a simple random sample of size s
# of a random variable X in {1, ..., n}
# with probability distribution p.
# Default values: s = 100, n = 10 and
# p = (p_1, ..., p_10), p_1 = ... = p_10 = 1/10.
{
muestra = sample(1:n, s, prob = p, replace = T)
fr = rep(0,n)
for (i in 1:n) fr[i] = sum(muestra = i)
fr = fr/s
frB = sum(fr[seq(2, n, 2)])
H.estimate = -sum( ifelse (fr > 0, fr*log2(fr), 0) )
H_B.estimate = -sum( ifelse (fr[seq(2,n,2)] >0,
fr[seq(2, n, 2)]/fr(B)*log2(fr[seq(2, n, 2)]/fr(B)), 0))
IB.estimate = -log2(sum(fr[seq(2, n, 2)]))
Suma = H_B.estimate + IB.estimate
error = abs(Suma-H.estimate)
return(cbind(H.estimate, H_B.estimate, IB.estimate, Suma, error))
}
```

APPENDIX B: ASSESSMENT TEST EXAMPLE

A source emits one of the symbols a, b, c, d, e, f with probabilities $p_1 = \frac{1}{2^5}$, $p_2 = \frac{1}{2^3}$, $p_3 = \frac{1}{24}$, $p_4 = \frac{1}{2}$, $p_5 = \frac{1}{22}$, $p_6 = \frac{1}{2^5}$, respectively.

A, B and C represent three people from whom we know that:

- A knows that the symbol b has been emitted.
- B knows that a vocal has been emitted.
- C does not know anything.

1. The potential information that C has is
   (a) 13/32
   (b) 31/16
   (c) 23/44
   (d) 15/32
2. The actual information of B is
   (a) $\log_2 7 + \log_2 3$
   (b) $\log_2 7$
   (c) 2.54
   (d) $5 - \log_2 9$
3. The difference between the uncertainty with respect to the symbol that the source has emitted that A and B have is
   (a) $\log_2 9 - \frac{8}{3}$
   (b) $\log_2 11$
   (c) $\log_2 9$
   (d) $\log_2 7 + 3$
4. We earn an amount of information less than the entropy of the source when we know that the emitted symbol was
   (a) a
   (b) b
   (c) c
   (d) d

5. Assume that a source emits the same symbols and has maximum entropy. Then, the value of p_5 and the maximum entropy are
   (a) $p_5 = 1, \quad H_{\max} = 6$
   (b) $p_5 = 1/2, \quad H_{\max} = \log_2 10$
   (c) $p_5 = 1/6, \quad H_{\max} = \log_2 6$
   (d) $p_5 = 1/6, \quad H_{\max} = \log_2 3$
6. If $p_i = 1/6$ for $i = 1, 2 \ldots 6$, then the sum of potential and actual information of B is
   (a) $1 + \log_2 3$
   (b) $\log_2 5$
   (c) 2
   (d) $\log_2 3 + \log_2 5$

Jorge Bonache is an associate professor in the Department of Applied Mathematics in the EUIT de Telecomunicacion at the Universidad Politecnica de Madrid, where he teaches undergraduate courses in mathematics. His current research focuses on probability, statistics and stochastic processes.

Carmen Cousido is an associate professor in the Department of Applied Mathematics in the EUIT de Telecomunicacion at the Universidad Politecnica de Madrid, where he teaches undergraduate courses in mathematics. His current research focuses on probability, statistics and stochastic processes.

Wilmar Hernandez is an associate professor in the Department of Circuits and Systems in the EUIT de Telecomunicacion at the Universidad Politecnica de Madrid, where he teaches undergraduate courses in analysis of circuits, and graduate courses in sensors and actuators. His current research focuses on assessment methodologies used to evaluate the effectiveness of educational experiments, and experiments involving courses, specific subjects and experiential activities.