How much information is contained in a signal?

Melvyn Tyloo and Kyle Wedgwood

FIG. 1. Telephone and telegraph wires in Broadway in 1890.

Information theory is ever more present in our daily life. One prominent example is when one sends a text message from one smartphone to another. The amount of memory needed to store the message, the channel capacity required to transmit it or the algorithm used to encode it, are all based on *information theory* [1]. These are just a few examples among the multitude of tasks performed in the background when using our electronic devices. All this has become possible thanks to the earlier development of information theory. In this project, we will dive into the elementary principles of information theory, first quantifying the information contained in a signal by modelling its uncertainty; second, when one is given two signals, we will measure the information that can be retrieved from one signal about the other.

Let us start be defining what is a *signal*. We call a signal a sequence of numbers $\{s_1, s_2, ..., s_N\}$ where $N \in \mathbb{N}$ is the length of the signal. In general, the elements of the sequence could be anything, e.g., measurements from an experiment, randomly generated numbers, output of a numerical simulation, any type of statistics. Here, we focus on signals that are made of natural numbers such that $s_i \in \mathbb{N}$ for $i = 1, ..., N$. We would like to measure the *uncertainty* of a signal. To have a better insight of what this means, let us imagine we were given the following fragment of signals made of 0's and 1's,

$$S_1 = \{0, 0, 0, 0, 0, 0, 0, 0\}\,,$$
$$S_2 = \{1, 1, 1, 1, 1, 1, 1, 1\}\,,$$
$$S_3 = \{1, 1, 0, 0, 0, 1, 1, 0\}\,.$$

Having a careful look at the signals, one would be rather confident in saying that the next number in S_1 might be

0 , and in S_2 it might be 1 . However, scrutinizing S_3 it is hard to infer any pattern that would allow one to predict the next number with some confidence. Intuitively, we would say that signal S_3 carries more uncertainty than S_1 and S_2 . Such intuitive reasoning can be formalised mathematically by means of the *entropy* of the signal.

Before moving on to the entropy, we first have to introduce the *empirical probability distribution* of a signal. The latter is an estimate of the probability that an element of the signal takes a particular value, which we denote $\mathbb{P}(s_i = \overline{s})$, $\overline{s} \in \mathbb{N}$. The empirical probability is obtained as,

$$\mathbb{P}_{\text{empirical}}(s_i = \overline{s}) = \frac{\text{number of times signal equals } \overline{s}}{\text{length of the signal}} \,.$$

By calculating the latter empirical probability for all the observed values in the signal, one obtains an estimate of the true probability distribution. Going back to the three signals $\{S_1, S_2, S_3\}$ given earlier, we calculate the empirical probability distributions,

$$\begin{aligned} S_1 &: \mathbb{P}_{\text{empirical}}(s_i = 0) = \frac{8}{8} = 1 \,; \quad \mathbb{P}_{\text{empirical}}(s_i = 1) = \frac{0}{8} = 0 \,, \\ S_2 &: \mathbb{P}_{\text{empirical}}(s_i = 0) = \frac{0}{8} = 0 \,; \quad \mathbb{P}_{\text{empirical}}(s_i = 1) = \frac{8}{8} = 0 \,, \\ S_3 &: \mathbb{P}_{\text{empirical}}(s_i = 0) = \frac{4}{8} = 0.5 \,; \quad \mathbb{P}_{\text{empirical}}(s_i = 1) = \frac{4}{8} = 0.5 \,. \end{aligned}$$

The empirical probability distribution for S_1 tells us that one obtains 0 with certainty (probability 1) and never obtains 1. The exact opposite is measured for S_2 where one obtains 1 with certainty and never obtains 0. For S_3 , one obtains 0 or 1 with equal probability (fifty-fifty chance). It is important to keep in mind that the empirical distribution is an estimate of the true probability distribution that becomes more and more accurate as the length of the signal increases.

Now that we know how to calculate the empirical probability distribution of a signal, let us define the entropy. Given a signal S with empirical probability distribution $\mathbb{P}(s_i = \overline{s})$, the entropy, H, of the signal is defined as,

$$H = -\sum_{j=1}^{M} \mathbb{P}(s_i = \overline{s}_j) \log[\mathbb{P}(s_i = \overline{s}_j)] \,,$$

where the sum runs over all the M possible values $\overline{s}_j$ the signal can take. The logarithm base is arbitrarily chosen depending on the unit. Here, we use the base two such that the unit is the bit. Entropy was introduced in the context of information theory by Claude Shannon in 1948. Let us calculate the entropy for a few signals to get more insights and see how the it relate to the uncertainty of the signals. For the above introduced signals one has,

$$\begin{aligned} S_1 &: H = -(1 \times \log(1) + 0 \times \log(0)) = 0 \,, \\ S_2 &: H = -(0 \times \log(0) + 1 \times \log(1)) = 0 \,, \\ S_3 &: H = -(0.5 \times \log(0.5) + 0.5 \times \log(0.5)) = 1 \,\text{bit} \,. \end{aligned}$$

One observes that the entropy for S_1 and S_2 is vanishing, while it is equal to one for S_3 . One also remembers that the uncertainty in predicting the next value for S_1 and S_2 is low, while there is a higher uncertainty for S_3 . Therefore, low values of the entropy indicate low uncertainty in the signal and high values mean that the signal is more uncertain. The entropy serves as a quantifier of the uncertainty of a signal. One can show that $0 \leq H \leq \log(M)$.

Exercise 1 *Calculate the empirical probability distribution and then the entropy of the following signals coming from one fair and one unfair coin:*

$$\begin{aligned} C_1 &= \{1, 1, 0, 0, 0, 0, 0, 1, 1, 0, 0, 1, 0, 1, 0, 0, 0, 1, 1, 1\} \,, \\ C_2 &= \{1, 0, 1, 1, 1, 0, 1, 0, 1, 0, 1, 1, 1, 1, 1, 1, 0, 1, 1, 0\} \,. \end{aligned}$$

Based on the entropy, identify the unfair coin.

Now that we know how to calculate the entropy of a signal, let us move to the more complicated situation where one would like to measure how much information is shared between two signals. This can be assessed using the *mutual information.* Let us consider the following two signals,

$$\begin{aligned} A &= \{a_1, a_2 ..., a_N\} \,, \\ B &= \{b_1, b_2 ..., b_N\} \,. \end{aligned}$$

These two signals can be analysed independently, but they can also be viewed in combination such that one has the joint signal,

$$C = \{(a_1, b_1), (a_2, b_2)..., (a_N, b_N)\}\,,$$

where the elements of the signal in C are made by taking the pairs (in order) of elements from A and B . Similarly to when analysing a single signal, one can calculate the *empirical joint probability distribution* for the signal C , which is the probability to observe a pair $(\overline{a}, \overline{b})$, given by,

$$\mathbb{P}_{\text{empirical}}[(a_i, b_i) = (\overline{a}, \overline{b})] = \frac{\text{number of times signals equals } (\overline{a}, \overline{b})}{\text{length of the joint signal}}\,.$$

Using the empirical probability distribution of each signal together with the joint probability distribution, we define the mutual information, I, as

$$I(A, B) = \sum_{j=1}^{M} \sum_{k=1}^{M} \mathbb{P}[(a_i, b_i) = (\overline{a}_j, \overline{b}_k)] \log \left(\frac{\mathbb{P}[(a_i, b_i) = (\overline{a}_j, \overline{b}_k)]}{\mathbb{P}(a_i = \overline{a}_j)\mathbb{P}(b_i = \overline{b}_k)} \right)\,.$$

To better understand mutual information, let us calculate it for the following signals. First, let us consider two independently randomly generated signals,

$$S_4 = \{1, 1, 1, 0, 0, 1, 1, 1, 0, 1, 0, 1, 1, 1, 1, 1, 1, 1, 1, 1\}\,,$$
$$S_5 = \{0, 0, 0, 0, 1, 1, 1, 0, 0, 0, 1, 1, 0, 0, 0, 1, 1, 0, 0, 1\}\,.$$

The joint signal reads,

$$S_{4,5} = \{(1{,}0), (1, 0), (1, 0), (0, 0), (0, 1), (1, 1), (1, 1), (1, 0), (0, 0), (1, 0),$$
$$(0, 1), (1, 1), (1, 0), (1, 0), (1, 0), (1, 1), (1, 1), (1, 0), (1, 0), (1, 1)\}\,.$$

The mutual information between signal S_4 and S_5 is given by $I(S_4, S_5) = 0.0074$ bit. As the two signals are independently generated, we expect the mutual information to be very close to zero. Let us consider now a third signal given by,

$$S_6 : \{0, 0, 0, 1, 1, 0, 0, 0, 1, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0\}\,.$$

The mutual information between S_4 and S_6 is given by $I(S_4, S_6) = 0.72$ bit, which is significantly different from zero. This means that S_4 and S_6 share a considerable amount of information. Examining closer the joint signal $S_{4,6}$, one notices that when one observes 1 (0) in S_4 , then one observes 0 (1) in S_6 . Such intrinsic connection translates into a finite mutual information.

Exercise 2 *Calculate the empirical joint probability distribution of the following signals:*

$$C_1 = \{1, 1, 0, 0, 0, 0, 0, 1, 1, 0, 0, 1, 0, 1, 0, 0, 0, 1, 1, 1\}\,,$$
$$C_2 = \{1, 0, 1, 1, 1, 0, 1, 0, 1, 0, 1, 1, 1, 1, 1, 1, 0, 1, 1, 0\}\,.$$

Calculate the mutual information $I(C_1, C_2)$.

Project *We want to study how information is transmitted in a simple model of a communication network. Let us consider a communication network given by Figure 2. An initial signal is created at node 1 and transmitted to nodes 2 and 3. The communication is made through so-called symmetric binary channels. This type of channel works as follows. Based on the original signal at node 1, a transmitted signal at node 2 is created where each element is a copy of the original element with probability ϵ, or the opposite [2] of the original element with probability $(1-\epsilon)$, where $0 \le \epsilon \le 1$. If $\epsilon = 1$, the channel transmits the original signal without any errors. Some level of error is expected whenever $\epsilon < 1$. We call ϵ, the transmission probability. Each channel in the communication network can have different probabilities ϵ. At the end of the communication network, node 4 receives two signals.*

1. *Generate a signal to transmit, and calculate its entropy.*

2. *Pick ϵ's for each of the channels. Generate the transmitted signals at node 2 and 3 and calculate their entropy.*

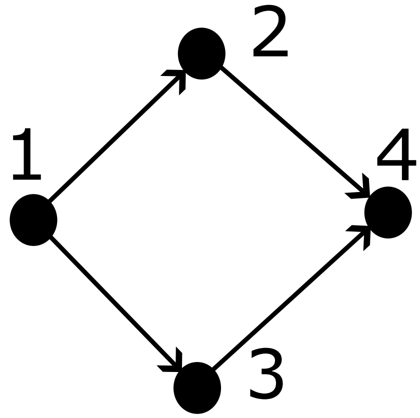

FIG. 2. Communication network.

3. *Generate the two transmitted signals at node 4 and calculate their entropy.*

4. *Compare the different entropies obtained along the communication network.*

5. *Calculate the mutual information between the signals at node 2 and 1, 3 and 1, 4 and 1.*

6. *Verify the Data-Processing inequality stating that the information related to the original signal can only decrease during such process. More specifically, if one has three signals X, Y, Z such that Y is obtained from X, and Z is obtained from Y, then the following inequality holds,*

$$I(X,Y) \geq I(X,Z)\,. \tag{1}$$

Going further *Communication networks can be very complex as depicted on Fig. 1. Imagine that you need to build a communication network that aims at maximizing the information transmission between a source and a target node. The total communication network is composed of 10 nodes, including the source and target ones. Apart from the source node, all nodes must have at least one incoming and one outgoing communication line. You can build two different types of communication line: one of higher quality with transmission probability $\epsilon_1 = 0.95$ and one of lower quality with $\epsilon_2 = 0.75$. The budget to create the network is limited such that one has access to 5 higher quality lines and 12 lower quality ones. These are the resources you have access to in order to build a communication network that maximises the information transmission between the source and the target nodes. When one node has multiple incoming lines, the effective message at the node is given by the most frequently observed value. If the two values are equally frequently received, the effective value is uniformly picked at random.*

After building your network, investigate the effect of a disruptive agent that introduces some additional noise at a single node (i.e., decreases $\epsilon_{1,2}$ at that node). Determine which node should the agent target to maximally disrupt the information transmission?

---

[1] T. M. Cover, *Elements of information theory* (John Wiley & Sons, 1999).
[2] The opposite of 0 is 1; the opposite of 1 is 0.