Société de Calcul Mathématique, S. A.
Algorithmes et Optimisation

About continuous and discrete entropy

Bernard Beauzamy

August 2007

As it is well-known, continuous and discrete entropy do not coincide, for a probability law.

Continuous entropy is defined as :

$$E_c = \int_{-\infty}^{+\infty} f(t) Log \frac{1}{f(t)}\, dt \tag{1}$$

where f is a density function ($f \geq 0$ and $\int f = 1$).

Discrete entropy is defined as :

$$E_d = \sum_i p_i Log \frac{1}{p_i} \tag{2}$$

where (p_i) is a sequence satisfying $p_i \geq 0$ and $\sum_i p_i = 1$.

Both represent a measure of the precision of the information : if the function f is very concentrated, E_c diminishes ; on the contrary, if the function f becomes flat, E_c increases. More precisely, if f tends to be concentrated at a single point (a Dirac measure), then $E_c \to -\infty$; take for instance the sequence of functions :

$$f_n(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp(-x^2 / 2\sigma^2) \text{ with } \sigma = 1/n \text{, when } n \to +\infty .$$

On the other hand, if f becomes flatter and flatter, then $E_c \to +\infty$; take for instance the sequence of functions :

$$f_n(x) = \frac{1}{2n} \text{ if } -n \leq x \leq n \text{, 0 otherwise.}$$

Similar properties are true with the discrete entropy E_d : it goes to $+\infty$ when the sequence p_i becomes less and less concentrated, but tends to 0 when this sequence concentrates : the lowest value is obtained when one of the p_i's is 1 and all the other are

0, and in this case $E_d = 0$. Also, quite obviously from the definition, $E_d \geq 0$ in all cases, which is not the case for E_c.

The problem with these definitions come from the following facts :

Take a continuous density f and discretize it : consider intervals I_j of equal size δ. Then define for all j :

$$p_j = \int_{I_j} f(x)\, dx \tag{3}$$

Then you can define both E_c from formula (1) and E_d from formula (2). We would expect that, when the intervals become smaller and smaller (that is, when $\delta \to 0$), E_d becomes closer and closer to E_c. But this is not true. In fact, $E_d \geq 0$,no matter what δ is, whereas E_c may very well be < 0. The precise relation is as follows :

Proposition. - When $\delta \to 0$,

$$E_c \approx E_d + Log(\delta)$$

Proof of Proposition.

Fix $\delta > 0$ and define each p_j by formula (3). Then :

$$E_d = \sum_j p_j Log\frac{1}{p_j} = \sum_j \left(\int_{I_j} f(x)\, dx \right) Log\frac{1}{\int_{I_j} f(x)\, dx} \quad (4)$$

Let, for each j, x_j be the center of the interval I_j. Then, when $\delta \to 0$,

$$\int_{I_j} f(x)\, dx \approx \delta f(x_j) \tag{5}$$

Substituting in (4), we get :

$$E_d \approx \sum_j \delta f(x_j) Log\frac{1}{\delta f(x_j)},$$

that is :

$$E_d \approx \delta \sum_j f(x_j) Log\frac{1}{f(x_j)} + \delta \sum_j f(x_j) Log\frac{1}{\delta} \tag{6}$$

But :

$$\delta \sum_j f(x_j) Log\frac{1}{f(x_j)} \approx \int_{-\infty}^{+\infty} f(x) Log\frac{1}{f(x)}\, dx \tag{7}$$

and :

$$\delta \sum_j f(x_j) \approx \int_{-\infty}^{+\infty} f(x)\, dx = 1 \tag{8}$$

So we get :

$$E_d \approx E_c + Log \frac{1}{\delta} \tag{9}$$

which proves our claim.

In practice, one should always work with the discrete entropy, because in real life phenomena only the discrete distribution (p_j) is known. This discretization corresponds to a loss of information. Assume that some phenomenon is highly concentrated (f is a Dirac measure). Then, if some interval of width δ is chosen and if the (p_j) are defined from formula (3), some information is lost : we know now that the phenomenon lies in one interval. So E_c can be close to $-\infty$ (very concentrated phenomenon) and still $E_d \geq 0$: they are linked by formula (9), which explains the difference between both.

Let us now turn to a concrete example

Example

Passing from continuous to discrete

We consider the density function :

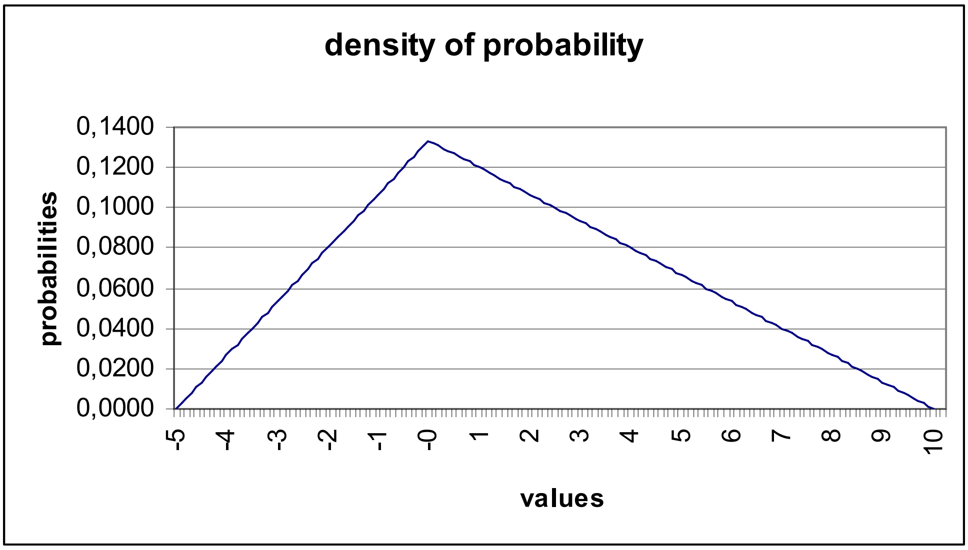
$y = 0$ if $x \leq -5$

$y = \frac{0.133}{5}(x+5)$ if $-5 \leq x \leq 0$

$y = -\frac{0.133}{10}(x-10)$ if $0 \leq x \leq 10$

$y = 0$ if $x \geq 10$.

Here is the graph of this density function :



Now, let us assume that we discretized it, with step $\delta = 0.1$ This means that we have the following probabilities (we give only the beginning of the list) , given by the formula $p_j = \delta f(x_j)$:

values	probas
-5,00	0,0000
-4,90	0,0003
-4,80	0,0005
-4,70	0,0008
-4,60	0,0011
-4,50	0,0013
-4,40	0,0016
-4,30	0,0019
-4,20	0,0021
-4,10	0,0024
-4,00	0,0027
-3,90	0,0029
-3,80	0,0032
-3,70	0,0035

The continuous entropy, computed for the function f above, using formula (1), gives :

$E_c = 2.5111$

The discrete entropy, computed from the list above, using formula (2), gives :

$E_d = 4.8173$

The corrected discrete entropy $E_{cd} = E_d + Log(\delta)$ is :

$E_{cd} = 2.5147$

Of course, E_c and E_{cd} are not exactly equal, because of the discretization process, but they are close, and they would be closer and closer when $\delta \to 0$.

Passing from discrete to continuous or step-wise constant

Assume now we start with the list of probabilities given in table above. Then, we can either reconstruct a continuous function (which, in our case, would be the function f defined above) or a step-wise constant function : given a probability p_j at a point x_j, the function g is defined as :

$$g(x) = \frac{p_j}{\delta} \text{ on the interval } I_j = \left[x_j - \frac{\delta}{2}, x_j + \frac{\delta}{2}\right]$$

So $\int_{I_j} g(x)dx = p_j$

The continuous entropy of g is :

$$\begin{aligned}
E_c(g) &= \int g(x) Log \frac{1}{g(x)} dx \\
&= \sum_j \int_{I_j} g(x) Log \frac{1}{g(x)} dx \\
&= \delta \sum_j \frac{p_j}{\delta} Log \frac{\delta}{p_j} \\
&= \sum_j p_j \left(Log\ \delta - Log\ p_j \right) \\
&= -\sum_j p_j Log\ p_j + Log\ \delta \\
&= E_d + Log\ \delta \\
&= E_{cd}
\end{aligned}$$

So we see that the corrected discrete entropy is equal to the continuous entropy of the step-wise constant function associated to the data. This step-wise constant function is simply the histogram associated with the data :