Société de Calcul Mathématique SA
Tools for decision help

Entropy and Information connected with a probability law

by Bernard Beauzamy

AMS Classification (2020) : 94A15, 94A17

Abstract

For a discrete probability law $p_1,...,p_N$, one may define its entropy and its variance; both characterize the "dispersion" of the law. We give comparison estimates between both and show that these estimates are best possible.

*****

Definition. - Let $p_1,...,p_N$ be a discrete probability law ($p_i \geq 0$, $\sum_{i=1}^{N} p_i = 1$). The entropy of this law is $I = -\sum_{i=1}^{N} p_i Log(p_i)$. This quantity, always positive, measures the "dispersion" of the law. Indeed, if the law is quite concentrated (all $p_i = 0$ except one equal to 1), then $I = 0$. Conversely, if the law has maximal dispersion (all $p_i = 1/N$), $I = Log(N)$, and this value is a maximum for I.

Link with Information Theory (see [Brilloin]). - If the numbers p_i are of the form $p_i = \frac{n_i}{N}$, where n_i is an integer, $I = -\sum_{i=1}^{N} p_i Log(n_i) + Log(N)$. The quantity $I_2 = \sum_{i=1}^{N} p_i Log_2(n_i)$ is the average (since we use the coefficients p_i) of the numbers $Log_2(n_i)$, which represent the number of necessary characters in order to write n_i in base 2. So we may consider I, after subtracting

$Log(N)$ and dividing by $Log(2)$, as an average of the number of characters needed in order to write each n_i.

Variance. - For a given sequence p_i as before, we define the variance of the sequence by the formula $V = \mathrm{var}(p_i) = \frac{1}{N}\sum_{i=1}^{N}\left(p_i - \frac{1}{N}\right)^2$. This definition is consistent with the usual definition, since the average of the p_i's is $\frac{1}{N}$. The variance is also a way to measure the concentration of the sequence, but of different nature. The minimum value of the variance is 0, attained when $p_i = \frac{1}{N}$ for all i. On the other hand,

$$V = \frac{1}{N}\sum_{i=1}^{N}\left(p_i - \frac{1}{N}\right)^2 = \frac{1}{N}\left(\sum p_i^2 - \frac{1}{N}\right) \le \frac{1}{N}\left(\sum p_i - \frac{1}{N}\right) = \frac{N-1}{N^2}$$

and the maximum value of the variance is attained when one p_i is equal to 1, all others 0. In fact:

- The variance measures the "geometric" dispersion of the values (considered for instance as points on an axis).
- The entropy measures the "probabilistic" dispersion of the values, that is the concentration of a probability law.

Their behavior is opposite: the variance is minimal when the entropy is maximal, and conversely. A better definition of the entropy might be the "corrected entropy", defined by:

$$I_c = \sum_{i=1}^{N} p_i Log(Np_i).$$

The corrected entropy has the same range of variation, but, for I_c, the variation goes in the same sense as the variance: they are both extreme at the same places. We now investigate the links between corrected entropy and variance.

Theorem 1. – *For any sequence* (p_i) *of length* N, *one has* $I_c \ge \alpha V$, with:

$$\alpha = \frac{N(N-1)}{N-2} Log(N-1).$$

This estimate is best possible.

Proof of Theorem 1

We need a Lemma:

Lemma 2. – *Let* (p_i) *be a sequence (with* $\sum p_i = 1$*) which realizes the minimum of* $I_c - \alpha V$. *Then this sequence takes at most two different values.*

Proof of Lemma 2

We may assume that the $p_i's$ are written in decreasing order, that is $p_1 \geq p_2 \geq \cdots \geq 0.$

Assume that $p_1 > p_2$. Take $\varepsilon > 0$ small enough and consider the sequence $(p_1 - \varepsilon, p_2 + \varepsilon, p_3, ..., p_N)$. Let $f(\varepsilon)$ be the quantity $I_c - \alpha V$ evaluated at this sequence. Then, $f(\varepsilon)$ is minimal for $\varepsilon = 0.$ We have $f'(\varepsilon) = Log\frac{p_2 + \varepsilon}{p_1 - \varepsilon} - \frac{2\alpha}{N}(p_2 - p_1 + 2\varepsilon)$, which leads to the condition $Log\frac{p_2}{p_1} = \frac{2\alpha}{N}(p_2 - p_1)$. Set $\beta = \frac{2\alpha}{N}$ and write $x = p_1 - p_2$. We obtain the equation $Log\left(1 + \frac{x}{p_2}\right) = \beta x$. A solution in x to this equation is obtained when one takes the intersection of the curve $y = Log\left(1 + \frac{x}{p_2}\right)$ with the straight line $y = \beta x$; it exists only if $p_2 < \frac{1}{\beta}$. If this condition is satisfied, then, for given $p_2,$ the equation has one and only one solution in x. Assume now that the minimal sequence (p_i) contains at least 3 different non-zero terms, that is $p_1 > p_2 > p_3 \geq p_4 \geq \cdots \geq p_N$.

Apply the above reasoning to the couples (p_1, p_3) and (p_2, p_3): the smallest one is the same in both cases. So, we deduce that the differences $p_1 - p_3$ and $p_2 - p_3$ must be the same. This shows that $p_1 = p_2$ and proves Lemma 2.

We now turn to the proof of Theorem 1. Assume, using Lemma 2, that n of the $p_i's$ take the value p and $N - n$ take the value q. Then $np + (N - n)q = 1,$ that is $q = \frac{1 - np}{N - n}$ and $0 \leq p \leq \frac{1}{n}.$

In this case, $I_c = npLog(Np) + (1 - np)Log\left(N\left(\frac{1 - np}{N - n}\right)\right)$ and:

$$V = \mathrm{var}(p_i) = \frac{1}{N}\sum_{i=1}^{N}\left(p_i - \frac{1}{N}\right)^2 = \frac{n}{N - n}\left(p - \frac{1}{N}\right)^2$$

We make a change of variables. We set $t = \frac{n}{N}$ and $x = np$, so $n = tN$, $p = \frac{x}{tN}$, and we have the intervals of variation $\frac{1}{N} \leq t \leq \frac{N - 1}{N}$, $0 \leq x \leq 1.$

The value $x = 0$ corresponds to $p = 0$; since $np + (N - n)q = 1,$ this means $q = \frac{1}{N - n}$. So, we have 0 repeated n times and $\frac{1}{N - n}$ repeated $N - n$ times. The value $x = 1$ corresponds to $np = 1,$ so $q = 0,$ and we have $\frac{1}{n}$ repeated n times and 0 repeated $N - n$ times. The extreme values $x = 0$ and $x = 1$ are possible.

In order to prove the Theorem, we need to show that, for all x and t:

$$x\,Log\frac{x}{t}+(1-x)Log\frac{1-x}{1-t}-\frac{(N-1)Log(N-1)}{(N-2)N\,t(1-t)}(x-t)^2\geq 0.$$

We set $\beta=\frac{(N-1)Log(N-1)}{(N-2)N}$ and $y=x\,Log\frac{x}{t}+(1-x)Log\frac{1-x}{1-t}-\beta\frac{(x-t)^2}{t(1-t)}$, considered as a function of x. We have $y'=Log\frac{x}{t}-Log\frac{1-x}{1-t}-2\beta\frac{x-t}{t(1-t)}$ and $y''=\frac{1}{x(1-x)}-\frac{2\beta}{t(1-t)}$.

We first observe that if $x=t,$ then $y=0$ and $y'=0$. The point $x=t$ is always a minimum for y ; however, depending on the value of $t,$ it may have another local minimum. In fact, three different shapes can be observed. We study the sign of y''.

Lemma 3.- Let t_N be the unique solution $<\frac{1}{2}$ of the equation $t(1-t)=\frac{(N-1)Log(N-1)}{2N(N-2)}$.

Then we have $\frac{1}{N}<t_N$ and, if $t_N\leq t\leq 1-t_N$, we have $y''>0$ for all x.

Proof of Lemma 3. - Let us first prove that $\frac{1}{N}<t_N$. Indeed, all we have to show is that $\frac{1}{N}\left(1-\frac{1}{N}\right)<\frac{(N-1)Log(N-1)}{2N(N-2)}$, which is equivalent to $\frac{N-2}{N}<\frac{Log(N-1)}{2}$. This is true for $N\geq 3$ and proves our claim.

Let us now prove the second statement. The condition $y''>0$ is equivalent to $x(1-x)<\frac{t(1-t)}{2\beta}$.

The maximum value of $x(1-x)$ is $1/4$. Therefore, the condition will be satisfied for all x if $\frac{t(1-t)}{2\beta}\geq\frac{1}{4}$, or $t(1-t)\geq\frac{(N-1)Log(N-1)}{2N(N-2)}$. This is equivalent to $t_N\leq t\leq 1-t_N$, and Lemma 3 is proved. A precise value of t_N is given by $t_N=\frac{1}{2}-\frac{1}{2}\sqrt{1-2\beta}$.

Proof of the Theorem in the case $t_N\leq t\leq 1-t_N$

In this case, we saw that $y''\geq 0$ for all x. So y' is increasing. But $y'\to-\infty$ when $x\to 0,$ $y'=0$ when $x=t,$ $y'\to+\infty$ when $x\to 1$. Therefore, the unique solution of $y'=0$ is obtained for $x=t$; $y'<0$ if $x<t$ and $y'>0$ if $x>t$. The minimum of y is obtained for $x=t$ and this minimum is 0. So $y\geq 0$ for all x and the Theorem is proved in this case. This is a simple case where the function y has only one minimum (namely $x=t$).

We now assume $\frac{1}{N} \le t \le t_N$; the discussion would be the same in the case $1-t_N \le t \le 1-\frac{1}{N}$, since all quantities are invariant under the transformations $t \to 1-t$ and $x \to 1-x$. First, we have to study the extreme case $t = \frac{1}{N}$.

Case $t = \frac{1}{N}$. We have: $y = x\,Log(Nx) + (1-x)\,Log\left(\frac{N(1-x)}{N-1}\right) - \frac{NLog(N-1)}{N-2}\left(x-\frac{1}{N}\right)^2$.

One verifies easily that y vanishes at the points $x = \frac{1}{N}$ and $x = 1-\frac{1}{N}$ and that y' vanishes at $x = \frac{1}{N}$, $x = \frac{1}{2}$ and $x = 1-\frac{1}{N}$. We now study the variations of y'. We know that $y'' \ge 0$ if and only if $x(1-x) \le \frac{t(1-t)}{2\beta} = \frac{\frac{1}{N}\left(1-\frac{1}{N}\right)}{2\beta}$. This happens if and only if $x \le x_N$ or $x \ge 1-x_N$, where x_N is the unique solution $< \frac{1}{2}$ of the equation $x_N(1-x_N) = \frac{\frac{1}{N}\left(1-\frac{1}{N}\right)}{2\beta}$. But $x_N > \frac{1}{N}$. So, we have the ordering $0 < \frac{1}{N} < x_N < \frac{1}{2} < 1-x_N < 1-\frac{1}{N} < 1$. The function y' is increasing between 0 and x_N, decreasing between x_N and $1-x_N$, increasing between $1-x_N$ and 1. The sign of y' is:

On $\left[0, \frac{1}{N}\right]$, $y' < 0$; on $\left[\frac{1}{N}, \frac{1}{2}\right]$, $y' > 0$; on $\left[\frac{1}{2}, 1-\frac{1}{N}\right]$, $y' < 0$; on $\left[1-\frac{1}{N}, 1\right]$, $y' > 0$.

So, the minimum is reached at the points $\frac{1}{N}$ and $1-\frac{1}{N}$ and the value of this minimum is 0.

This proves that $y \ge 0$ for all x, when $t = 1-\frac{1}{N}$, and finishes the proof of the Theorem in the case $t = 1-\frac{1}{N}$.

Case $\frac{1}{N} < t \le t_N$. There is some difference with the case $t = \frac{1}{N}$. There will be a second local minimum (besides $x = t$), but the value at this second minimum will not be 0. The function y does not vanish between $1-\frac{1}{N}$ and 1, but still has a minimum in this interval.

We saw that the condition $y'' > 0$ is equivalent to $x(1-x) < \frac{t(1-t)}{2\beta}$ and we are in the case $\frac{t(1-t)}{2\beta} < \frac{1}{4}$. Therefore, $y'' > 0$ if $x < x_t$ and if $x > 1-x_t$, where x_t is the unique solution $< 1/2$ of the equation $x(1-x) = \frac{t(1-t)}{2\beta}$. We observe that $t < x_t$. Indeed, x_t is solution of this equation,

since $2\beta < 1$. So, we have $0 < t < x_t < \frac{1}{2}$. The proof will follow different patterns, depending on the position of x.

We have $y'' > 0$ if $0 \le x \le x_t$, $y'' = 0$ for $x = x_t$, and $y'' < 0$ if $x_t \le x \le 1 - x_t$. Therefore, y' is increasing if $0 \le x \le x_t$, decreasing if $x_t \le x \le 1 - x_t$ and increasing if $1 - x_t \le x \le 1$. At the point 0, y' has limit $-\infty$ and y' vanishes at $x = t$; therefore $y' < 0$ if $0 < x < t$. So y is decreasing on this interval.

Also, y' is increasing between t and x_t and therefore is positive, since y' vanishes at $x = t$. Therefore, y is increasing between t and x_t.

Case $0 \le x \le x_t$ - Then, the function y is first decreasing, then increasing; it reaches its minimum at $x = t$ and this minimum is 0. So, $y \ge 0$ on this interval.

Case $x_t \le x \le \frac{1}{2}$.- On the interval $x_t \le x \le 1$, y' is first decreasing (if $x_t \le x \le 1 - x_t$), reaches its minimum at $1 - x_t$, then is increasing if $1 - x_t \le x \le 1$. We know that $y'(x_t) > 0$ and that the limit of y' at $+\infty$ is $+\infty$. If $y'(1 - x_t) < 0$, y' vanishes twice, is positive first, then negative, then positive; y is first increasing, then decreasing, then increasing, and there is a local minimum of y between x_t and 1. We will first show that $y' > 0$ when $x = \frac{1}{2}$. We have:

$$y'(1/2) = Log\left(\frac{1}{2t}\right) - Log\left(\frac{1}{2(1-t)}\right) - 2\beta\frac{\frac{1}{2} - t}{t(1-t)} = Log\frac{1-t}{t} - \beta\frac{1-2t}{t(1-t)}$$

We set $A = Log\frac{1-t}{t} - \beta\frac{1-2t}{t(1-t)}$ and we have:

$$A' = -\frac{1}{t(1-t)} - \beta\frac{-2t^2 + 2t - 1}{t^2(1-t)^2} = \frac{-t(1-t) - \beta(-2t^2 + 2t - 1)}{t^2(1-t)^2}$$

which has the same sign as:

$$B = -t(1-t) - \beta(-2t^2 + 2t - 1) = t^2(1 + 2\beta) - t(1 + 2\beta) + \beta$$

This is a quadratic function, which reaches its absolute minimum at $t = \frac{1}{2}$ and so it is decreasing for $t < \frac{1}{2}$. Since $t < t_N$, the minimum is obtained for $t = t_N$; its value is:

$$C = t_N^2(1 + 2\beta) - t_N(1 + 2\beta) + \beta = (1 + 2\beta)(t_N^2 - t_N) + \beta$$

But, by definition, $t_N\left(1-t_N\right)=\frac{\beta}{2}$. So, $C=-\left(1+2\beta\right)\frac{\beta}{2}+\beta=-\beta^2+\frac{\beta}{2}=\beta\left(\frac{1}{2}-\beta\right)>0$.

This shows that $B>0$, so that $A'>0$ for all t; therefore A is increasing. The smallest value is obtained at $t=\frac{1}{N}$: $A=Log\left(N-1\right)-\beta\frac{N\left(N-2\right)}{N-1}=0$, which proves that $A\geq 0$ if $\frac{1}{N}\leq t\leq t_N$, when $x=\frac{1}{2}$. We will now show that $y'>0$ on the interval $t<x<1/2$. We have $y'=Log\frac{x}{1-x}\frac{1-t}{t}-2\beta\frac{x-t}{t\left(1-t\right)}$. We set $A=Log\frac{x}{1-x}\frac{1-t}{t}-2\beta\frac{x-t}{t\left(1-t\right)}$, considered as a function of t. We have: $A'=-\frac{1}{t\left(1-t\right)}-2\beta\frac{-t^2+2xt-x}{t^2\left(1-t\right)^2}=\frac{-t\left(1-t\right)-2\beta\left(-t^2+2xt-x\right)}{t^2\left(1-t\right)^2}$, which has same sign as: $B=-t\left(1-t\right)-2\beta\left(-t^2+2xt-x\right)=t^2\left(1+2\beta\right)-t\left(1+4\beta x\right)+2\beta x$.

But, considered as a function of x, this last quantity is decreasing, therefore takes its minimum for $x=1/2$; we are back to the previous case, and we have shown that $B>0$, so that $A'>0$, and therefore that A is increasing as a function of t, for fixed x.

Since $A=0$ if $t=x$, we have $A>0$ if $t<x\leq\frac{1}{2}$, which shows that $y'>0$ on this interval. This implies that, as a function of x, y is increasing if $t<x<\frac{1}{2}$. Since $y=0$ if $x=t$, we have $y>0$ if $t<x<\frac{1}{2}$ and the Theorem is proved on the interval $0<x<\frac{1}{2}$.

Case $\frac{1}{2}<x\leq 1$. - Let $y=xLog\left(\frac{x}{t}\right)+\left(1-x\right)Log\left(\frac{1-x}{1-t}\right)-\beta\frac{\left(x-t\right)^2}{t\left(x-t\right)}$. We have:

$$\frac{\partial y}{\partial t}=\frac{-x}{t}+\frac{1-x}{1-t}-\beta\frac{\left(x-t\right)\left(2xt-t-x\right)}{t^2\left(1-t\right)^2}=\frac{\left(t-x\right)t\left(1-t\right)-\beta\left(x-t\right)\left(2xt-t-x\right)}{t^2\left(1-t\right)^2}$$

Since $x-t>0$, this quantity has same sign as:

$$A=-t\left(1-t\right)-\beta\left(x-t\right)\left(2xt-t-x\right)=t^2-2\beta tx+\beta t-t+\beta x$$

We will show that $A\geq 0$ for all x and all t. We have: $\frac{\partial A}{\partial x}=-2\beta t+\beta=\beta\left(1-2t\right)>0$.

So A is an increasing function of x and the minimum is obtained for x minimum, which is $x=\frac{1}{2}$. For this value, we have $A_{\min}=t^2-t+\frac{\beta}{2}$ and $\frac{\partial A_{\min}}{\partial t}=2t-1<0$.

So $A_{\min}$ is a decreasing function of t and the minimum of $A_{\min}$ is obtained for t maximal, that is for $t=t_N$. But t_N is defined by the equation $t_N\left(1-t_N\right)=\frac{\beta}{2}$ and therefore $A_{\min}=0$, which

proves that $A \geq 0$. This implies that y is an increasing function of t, so the minimum of y is obtained for t as small as possible, that is for $t = t_N$. But, for this value of t, we know that $y \geq 0$ for all x and the Theorem is proved.

The fact that the estimate in Theorem 1 is best possible follows from the sequence :

$$p_1 = 1 - \frac{1}{N}, p_2 = ... = p_N = \frac{1}{N(N-1)}$$

In this case $I_c = \frac{N-2}{N} Log(N-1)$, $V = \frac{1}{N-1}\left(\frac{N-2}{N}\right)^2$, and $\frac{I_c}{V} = \alpha$.

We now turn to converse estimates, relating I_c and the variance. We have:

Theorem 4. – *For any sequence* (p_i) *of length* N, *we have* $I_c \leq N^2 V$ and this estimate is best possible.

Proof of Theorem 4. - We have $Log(Np_i) \leq Np_i - 1 = N\left(p_i - \frac{1}{N}\right)$ and:

$\sum p_i Log(Np_i) \leq N \sum p_i \left(p_i - \frac{1}{N}\right) \leq N \sum \left(p_i - \frac{1}{N}\right)^2$, which proves Theorem 4.

The constant N^2 is best possible, as the following example shows:

$n = 1$ $p_1 = \frac{1}{N^2}$, $p_2 = ... = p_N = \frac{1}{N} + \frac{1}{N^2}$. Then $\frac{I_c}{V} \sim N^2$ when $N \to +\infty$.

We deduce from the previous Theorems:

Corollary 5. – *For the entropy* I, *we have the following bounds, which are best possible:*

$$Log(N) - N^2 V \leq I \leq Log(N) - \frac{N(N-1)Log(N-1)}{N-2} V$$

Reference

[Brilloin] Léon Brillouin La science et la théorie de l'information, 1959, réédité par les Editions Jacques Gabay, Paris.