MaxEnt 2010   Chamonix

The logarithm ubiquity in statistics, information, acoustics …

Michel Maurin

table

1 - recalls

Logarithm function is a very well used and known function in mathematics, the first tables by Napier, Briggs, … , the Leibniz formula (Newton, Bernoulli ?) roughly sixty years later,

it has many properties, fine and suitable, and it is also very widely used in other scientific domains, for instance

* the information in probability, Hartley - log P(A)

Shannon - $\log_2$ P(a)

* the entropy in physics, Clausius and $k \ln W$ of Boltzmann,

in probability, Shannon E (- log $P(A_i)$) for a $P(A_i)$ distribution

* the information in Statistics : when $P_\theta(A_i)$ depending on θ,

$\sum_i \; d\, P_\theta(A_i)/d\theta = 0$, and then $\sum \; P_\theta(A_i) \; d \ln P_\theta(A_i)/d\theta = 0$,

the $d \ln P_\theta(A_i)/d\theta$ distribution is centered ; the Fisher information is its variance

$$\sum_i \; P_\theta(A_i) \, [d \ln P_\theta(A_i)/d\theta]^2 \quad ;$$

and with two distributions P_1 and P_2 the Kullback information is the expectancy of the log ratio $P_2(A_i)/ P_1(A_i)$,

$$\sum_i \; P_1(A_i) \ln P_2(A_i)/ P_1(A_i) \quad ;$$

* in chemistry we get the pH = - log $[H^+]$;

* in seismology with Richter scale ;

* the special case of acoustics

in this domain one is commonly dealing with a "level" of magnitude g, say

$$L = 10 \log g/g_0$$

g_0 for conventional reference, in decibels ;

"quite every" positive thing may be transformed in level …,

with not quite clear reasons in order to this definition. The calculations with noise levels has to obey to a "logic of levels", of course not usual nor arithmetic,

* and econometry with its elasticity coefficient $e = \dfrac{dx/x}{dy/y}$,
and then with implicit levels for x and y.

2 - We introduce a "new" statistic in order to describe the dispersion of data x_i, and there is a special role for logarithm end exponential..

A - let the h-mean $h^{-1}(1/n \sum h(x_i))$;

we have $h^{-1}(1/n \sum h(x_i)) > m_x = 1/n \sum x_i$ whenever h is an increasing and convex function,

then $h^{-1}(1/n \sum h(x_i)) - m_x > 0$ as a variance ;

we note the difference $T_2 = h^{-1}(1/n \sum h(x_i)) - m_x$, positive.

B - as a second common property with variance we suppose T_2 is only depending on differences $\Delta x_i = x_i - m_x$. as do the variance $\sigma^2 = 1/n \sum \Delta x_i^2$

this property is due to a N S C theorem (if and only if) :

one has $T_2 = S(\Delta x_i)$ if and only if $h(x) = e^{cx}$, $c > 0$;

a) it is sufficient,

b) for the necessity the proof comes from functional equations' method (Aczél), and then $T_2 = 1/c \ln\{1/n \sum e^{c \Delta x_i}\}$

it is called the « expo-dispersion », in place of a previous h-dispersion

T_2 , or $T_2(\mathbf{X})$, $T_2(x_i)$ has an important lot of properties, analytical and also statistical.

$T_2(\mathbf{X}+\mathbf{Y}) = T_2(\mathbf{X}) + T_2(\mathbf{Y})$, with independency of random variables

$T_2(x_i)$ in invariant for a d_x translation on x variable, as Δx_i ,

$T_2(\Delta x_i)$, as depending on $\sum x_i$ is convex in Δx_i ,

$T_2(\Delta x_i)$ is increasing in c ,

let $x_i = m + \Delta x_i$ initial data, and $x_{i,t} = m_x + t\,\Delta x_i$, $t > 0$, $x_i = x_{i,1}$, and $\kappa(t) = T_2(x_{i,t}) = T_2(m_x, t\Delta x_i)$; then

$\kappa(t)$ is an increasing monotony in t for every increasing convex h,

$\kappa(t)$ is a convex function in t with $h(x) = e^{cx}$, $c > 0$;

- application to acoustics :

$L_i = 10 \log p_i^2/p_0^2$ (the conventional definition for level)

with p_i the acoustic pressure (sur-pressure), power p_i^2 additive (when sounds are independent), and p_0 a reference. And (remind) one uniquely deals with noise levels L_i.

Then $10^{Li/10} = p_i^2/p_0^2 = e^{Li/M}$ are additive, ($M = 10/\ln 10$),

the expo-dispersion for noise levels L_i is, when taking $c = 1/M$,

$$T_2(L_i) = M \ln[1/n \sum e^{(Li-Lm)/M}] = 10 \log[1/n \sum 10^{Li/10}\ 10^{-Lm/10}]$$

$$= 10 \log[1/n \sum 10^{Li/10}] - L_{arith\text{-}mean} = L_{eq} - L_{arith\text{-}mean} ,$$

a suitable dispersion index for noise levels into the logic of levels.

3 - the most fine statistical property, (finest ?)

when mixing sub-populations x_{ij} one has (classic)

final **var** = mean of $\boldsymbol{\sigma_j^2}$ + **var** of mean_j^2

and here

final **expo-dispersion** = mean of $\mathbf{T_{2j}}$ +

expo-dispersion of h-means j

and also with entropy, (MaxEnt 2000)

final **entropy** = mean of $\mathbf{entropy_j}$ +

entropy of the weights j of sub-populations then one has three statistics for describing the dispersion of data, variance, entropy and expo-dispersion,

- variance and expo-dispersion for quantitative data,
- entropy for qualitative ones,

and they share the same structural decomposition when there are sub-populations, and a mixing of them.

Then as a conclusion, this is the story of statistic, T_2, built with the exponential function,which is convenient and suitable for acoustics and its specific levels, and sharing a fine property with entropy.

And all of this is due to the logarithm and its many (ubiquitous) properties, this famous *logos* about *arithmos* so named by John Napier, who also declared about them *mirifici logarithmorum canonis descriptio* …(here in latin).

some references

- Une formule de décomposition de l'entropie, une interprétation relative au mélange de lois, **MaxEnt 2000**, Gif sur Yvette

- Un amusement sur la dispersion des données, la h-dispersion, **XXXIXe Journées SFdS 2007**, Angers, version de février 2009

- Acoustique, statistique et “logique des niveaux”, **10ème CFA, 2010**, Lyon

- Sur des compromis entre la logique des niveaux et les statistiques classiques de niveaux de bruit, **INRETS 2008**

- Logarithme, niveaux, décibels et "logique des niveaux", **INRETS 2006**

Thank you

for your

attention