Point information gain, point divergence gain and respective entropies and entropy densities as measures of information of multidimensional discrete phenomena

Dalibor Štys, Renata Rychtáriková

May 22, 2018

State of the Art: Entropy

Measurement of relative information between two probability distributions:

a) Shannon's information entropy

b) Kullback-Leibler divergence

$$D_{KL}(P||Q) = \sum_i p_i \ln \frac{p_i}{q_i} = E_p\left[\ln q_i\right] - E_p\left[\ln p_i\right] = S_P(Q) - S(P), \tag{1}$$

where $S_P(Q)$ is so-called cross-entropy and $S(P)$ is the entropy of distribution P. In case, when P is similar to Q, this measure can be approximated by entropy difference

$$\Delta S(P,Q) = S(Q) - S(P). \tag{2}$$

c) Rényi's information entropy

$$\mathscr{H}_\alpha = \frac{1}{1-\alpha} \ln\left(\sum_{j=1}^{k} p_j^\alpha\right) \tag{3}$$

d) Generalised dimension in multifractal systems is based on Rényi entropy

$$D_\alpha = \lim_{l \to 0} \frac{\mathscr{H}_\alpha(P(l))}{\ln l} \tag{4}$$

e) In imaging using lenses the resulting interference pattern is multifractal because in each point we have contribution of all imaged objects, their part in focus as well as outside focus

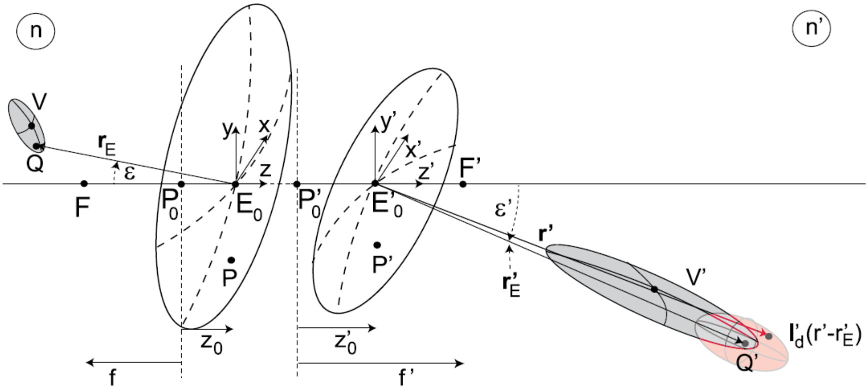


Figure 1: *Image of a volume objects from Braat J.J.M., Janssen A.J.E.M., Derivation of various transfer functions of ideal or aberrated imaging systems from the three-dimensional transfer function, JOSA A, 2015*

Our approach for obtaining the definition of information measure

Task:

What is the contribution of each element to the total information of the multidimensional discrete space?

Solutions:

1) Point information gain (PIG, $\Gamma_{\alpha,i}$):

$$\Gamma_{\alpha,i} = \mathscr{H}_{\alpha,i} - \mathscr{H}_{\alpha} = \frac{1}{1-\alpha}\ln\left(\sum_{j=1}^{k} p_{j,i}^{\alpha}\right) - \frac{1}{1-\alpha}\ln\left(\sum_{j=1}^{k} p_{j}^{\alpha}\right), \quad (5)$$

It may be written as

$$\Gamma_{\alpha,i} = \frac{1}{1-\alpha}\ln\left(\frac{\sum_{j=1}^{k} p_{j,i}^{\alpha}}{\sum_{j=1}^{k} p_{j}^{\alpha}}\right) \quad (6)$$

where α is the Rényi coefficient, k is the number of phenomena in the discrete distribution, $p_j = n_j/n$ and $p_{j,i} = n_{j,i}/(n-1)$ is the probability of occurrence of the j-th phenomenon in the original distribution and in the distribution without one element i of the j-th phenomenon, respectively.

2) Point information gain entropy (PIE, H_α):

$$H_\alpha = \sum_{j=1}^{k} n_j \Gamma_{\alpha,j}. \tag{7}$$

3) Point information gain entropy density (PIED, Ξ_α):

$$\Xi_\alpha = \sum_{j=1}^{k} \Gamma_{\alpha,j}. \tag{8}$$

Properties of $\Gamma_{\alpha,i}$, H_α, and Ξ_α

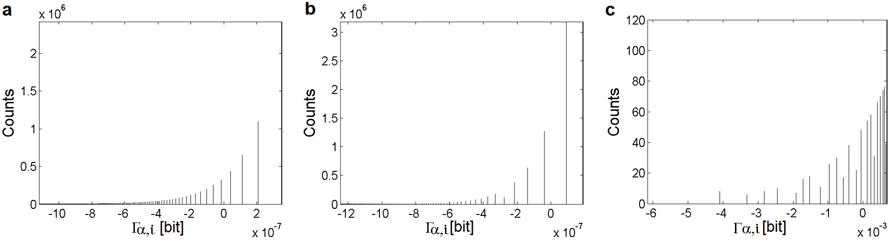


Figure 2: $\Gamma_{\alpha,i}$*-transformations of the discretized Lévy (**a**), Cauchy (**b**), and Gauss (**c**) distribution at* $\alpha =$ *0.99. The deviation from the monotonous dependency in the Gauss distribution is due to the digital rounding.*

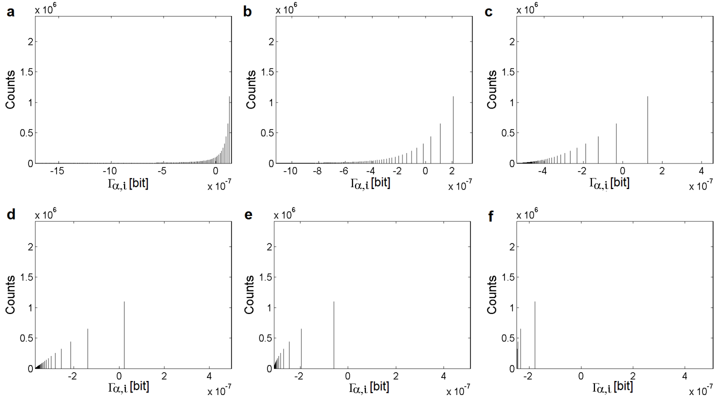


Figure 3: $\Gamma_{\alpha,i}$*-transformations of the discretized Lévy distribution at* $\alpha = \{0.5; 0.99; 1.5; 2.0; 2.5; 4.0\}$ *(from* ***a*** *to* ***f****). The dependency of* $\Gamma_{\alpha,i}$ *on number of counts is nearly linear for* $\alpha = 2$.

It may be shown using Taylor expansion that

$$\Gamma_{2,i} \approx \mathcal{C}_2(n) + \frac{1}{1-2}\left(\frac{\sum_{j=1,j\neq i}^{k-1} n_j^2 + (n_i-1)^2}{\sum_{j=1}^{k} n_j^2} - 1\right) \approx \mathcal{C}_2(n) + \frac{2n_i - 1}{\sum_{j=1}^{k} n_j^2}, \qquad (9)$$

Application of $\Gamma_{\alpha,i}$, H_α, and Ξ_α to image processing and analysis

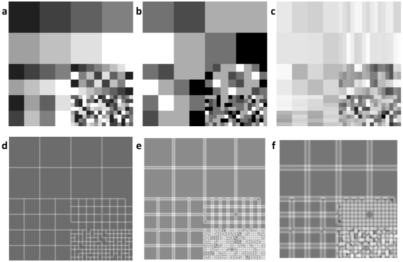


Figure 4: *Standard images.* $\Gamma_{0.99,i}$*-transformations of the texmos2.s512 image. Original image (**a**) and information images calculated from the whole image (**b**), a cross around each pixel (**c**), and squares of the side of 5, 15, and 29 px, respectively, with the centered examined pixel(**d–f**).*

Fakulta rybářství a ochrany vod
Faculty of Fisheries and Protection of Waters
Jihočeská univerzita v Českých Budějovicích
University of South Bohemia in České Budějovice

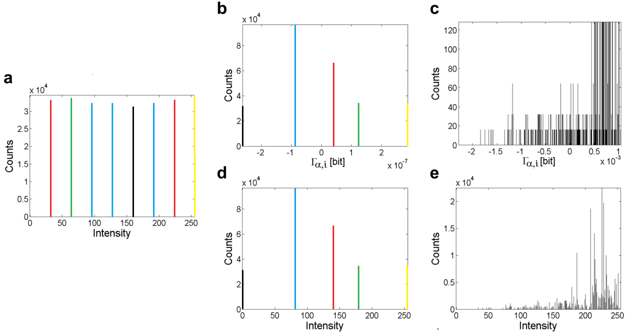


Figure 5: *Standard images. Histograms of $\Gamma_{0.99,i}$-transformations of the texmos2.s512 image. Original image (**a**), original $\Gamma_{0.99,i}$-values calculated from the whole image (**b**), original $\Gamma_{0.99,i}$-values calculated from a cross whose shanks intersect in the examined pixel (**c**), $\Gamma_{0.99,i}$-transformed images calculated from whole image (**d**), and $\Gamma_{0.99,i}$-transformed images calculated from a cross around each pixel (**e**). Colors in the original and globally (whole image) transformed histograms correspond to the intensity levels with the identical frequencies of occurrences in the original image.*

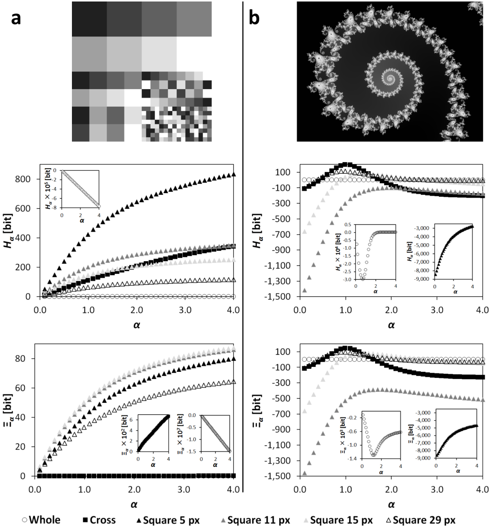
Point information gain entropy (PIE, H_α)

$$H_\alpha = \sum_{j=1}^{k} n_j \Gamma_{\alpha,j} \tag{10}$$

Point information gain entropy density (PIED, Ξ_α)

$$\Xi_\alpha = \sum_{j=1}^{k} \Gamma_{\alpha,j} \tag{11}$$

Figure 6: *Standard images.* H_α- and Ξ_α-spectra for semantic information and different syntactic surroundings of a unifractal (texmos2.s512, column **a**) and multifracal (wd950112, column **b**) image at $\alpha = \{0.1, 0.2, ..., 0.9, 0.99, 1.1, 1.2, ..., 4.0\}$. The maximum in the mulstifractal image is in all cases at the position of $\alpha = 1$ and the inflection point at $\alpha = 2$.



State of the Art: Divergence

Measurement of relative information between two probability distributions:

The Point Divergence Gain (PDG, $\Omega_\alpha^{(l \to m)}$), where a discrete distribution $P^{(i)}$ is replaced by a distribution

$$P^{(l \to m)} = \left\{ p_j^{(l \to m)} \right\}_{j=1}^{k} = \left\{ \frac{n_1}{n}, \ldots, \frac{n_l - 1}{n}, \ldots, \frac{n_m + 1}{n}, \ldots, \frac{n_k}{n} \right\}, \qquad (12)$$

which can be obtained from the original distribution P, where the occurrence of the examined l-th phenomenon ($n_l \in \mathcal{N}^+$) is removed and supplied by a point of the occurrence of the m-th phenomenon ($n_m \in \mathcal{N}_0$).

$\Omega_\alpha^{(l\to m)}$ does not depend (contrary to the $\Gamma_\alpha^{(i)}$) on n but depends only on the number of elements of each phenomenon j. Let us design the nominator $\sum_{j=1}^{k} n_j^\alpha$, which is constant and related to the original distribution (histogram) of elements and to the parameter α, as $\mathcal{C}_\alpha$. It gives us the final form

$$\Omega_\alpha^{(l\to m)} = \frac{1}{1-\alpha} \ln\left[\frac{(n_l-1)^\alpha - n_l^\alpha + (n_m+1)^\alpha - n_m^\alpha}{\mathcal{C}_\alpha} + 1\right]. \tag{13}$$

For a particular distribution, $\Omega_\alpha^{(l\to m)}$ is a function only of the parameter α and frequencies of occurrences of the phenomena n_l and n_m in the original distribution, between which the exchange of the element occurs.

Now we will consider the specific case $\alpha = 2$ (collision entropy) for which it may be written

$$\Omega_2^{(l\to m)} = -\ln\left[\frac{2}{\mathcal{C}_2}(n_m - n_l + 1) + 1\right] = -\ln\left[\frac{2}{\mathcal{C}_2}(\Delta n^{(l\to m)} + 1) + 1\right]. \tag{14}$$

For a specific difference $\Delta n^{(x\to y)} = D$, which can be approximated by the 1st-order Taylor sequence

$$\begin{aligned}\Omega_2^{(l\to m)} &\approx -\ln\left[\frac{2}{\mathcal{C}_2}(D+1)+1\right] - \frac{2}{2(D+1)+\mathcal{C}_2}(\Delta n^{(l\to m)} - D)\\ &= -\frac{2}{2D+2+\mathcal{C}_2}\Delta n^{(l\to m)} + \frac{2D}{2D+2+\mathcal{C}_2} - \ln\left[\frac{2D}{\mathcal{C}_2}+\mathcal{C}_2+1\right].\end{aligned} \tag{15}$$

Which shows the connection between Ω_α and simple subtraction. It will be shown in the next presentation that in correct datasets Ω_α may be successfully replaced by simple subtraction, i.e. multifractality may be neglected.

Point Divergence Gain Entropy and Point Divergence Gain Entropy Density

The resulting $\Omega_{\alpha}^{(a_i \to b_i)}$ then quantifies how much information is gained/lost, when, at the i-th position, we replace the value a_i for the value b_i. A Point Divergence Gain Entropy (PDGE, I_α) is defined as a sum of absolute values of all PDGs for all pixels, i.e.,

$$I_\alpha(\mathcal{I}_a; \mathcal{I}_b) = \sum_{i=1}^{n} |\Omega_{\alpha}^{(a_i \to b_i)}| = \sum_{l=1}^{k} \sum_{m=1}^{k} n_{lm} |\Omega_{\alpha}^{(l \to m)}|, \tag{16}$$

where n_{lm} denotes the number of present substitutions $l \to m$, when we transform $\mathcal{I}_a \to \mathcal{I}_b$. The absolute value ensures that the contribution of the transformation of a rare point to a frequent point (negative Ω_α) and a frequent point to a rare point (positive Ω_α) do not cancel each other and both contribute to the resulting PDGE.The PDGE can be understood as an absolute information change.

Fakulta rybářství a ochrany vod
Faculty of Fisheries and Protection of Waters
Jihočeská univerzita v Českých Budějovicích
University of South Bohemia in České Budějovice

Moreover, it is possible to introduce other macroscopic quantity – a Point Divergence Gain Entropy Density (PDGED, P_α), where we do not sum over all pixels, but only over all realized transitions $l \to m$. Thus, the PDGED can be defined as

$$P_\alpha(\mathcal{I}_a;\mathcal{I}_b) = \sum_{l=1}^{k}\sum_{m=1}^{k} \chi_{lm}|\Omega_\alpha^{(l\to m)}|, \tag{17}$$

where

$$\chi_{lm} = \begin{cases} 1, & n_{lm} \geq 1; \\ 0, & n_{lm} = 0. \end{cases} \tag{18}$$

We can understand the quantity PDGED as an absolute information change of all realized transitions of phenomena $m \to l$.

Applications

Each multidimensional dataset is characterised by a vector of Γ_α and / or Ξ_α values. If we know the appropriate surrounding, we may use it for the calculation of the Γ_α or Ξ_α value. If not, calculation from the whole image means that we ignore the structure of the system while calculation using the cross method comprises (blindly) all possible contexts. The resulting values are typically used as initial variables in multivariate analysis. Weights (contributions) of each principle component may be understood as spectral contribution of "measurable" variables to the combined "internal" variable.

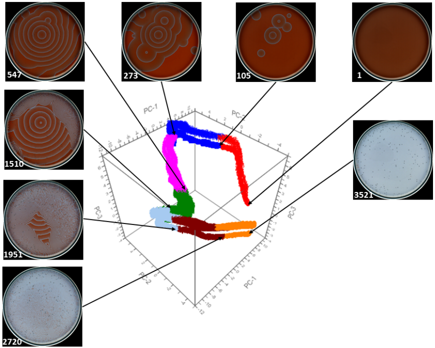


Figure 7: ***Left panel:*** *Trajectory in principal component space of series of images of the chemical self-organisation (Belousov-Zhabotinsky reaction) based on combination of whole and cross surroundings for 13* Γ_α *values for each colour channel. (2x39 values).*

In each physico-chemical textbook is introduced the idea of the chemical potential μ_i and the activity a_i which are the real measures of the contribution of each molecule to total Gibbs energy G of the examined system. It is not from the first sight controversial to write the total Gibbs energy as

$$G = \sum_{i=0}^{n} \mu_i, \tag{19}$$

where index i determines the individual chemical component of n components present in the mixture. We should, however, expand μ_i as

$$\begin{aligned} \mu_i = \mu_{0,i} + \nu_i RT \ln(a_i) = \mu_{0,i} + \nu_i RT \ln(c_i * \gamma_i) = \mu_{0,i} + \\ \nu_i RT \ln(c_i) + \nu_i RT \ln(f(c_1, c_2 ... c_n, T, p, V ...), \end{aligned} \tag{20}$$

where $\mu_{0,i}$ is the standard chemical potential of the i-th component of concentration c_i, $\gamma_i = f(c_1, c_2 ... c_n, T, p, V ...)$ is its activity coefficient, ν_i is the respective stoichiometric coefficient, and R is the universal gas constant.

The activity coefficient is in principle a function of concentrations c_i of all components in the mixture and all other relevant state variables such as temperature T, pressure p, volume V, etc. The replacement $\gamma_i = 1$ by c_i is impossible, each compound must be accounted together with its interactions with its surrounding. Γ_i may be understood as the equivalent of μ_i and Ξ_i as the equivalent of γ_i in structured system. In fact, however, the intermolecular interactions have both asymmetry (= geometric directionality) and distance dependency and are from certain point of view multifractal too.

The generalized **Point Divergence Gain** $\Omega_{\alpha}^{(l \to m)}$ was originally used for characterization of dynamic changes in image series, namely in z-stacks of raw RGB data of unmodified live cells obtained via scanning along the z-axis using video-enhanced digital bright-field transmission microscopy.

Fakulta rybářství a ochrany vod
Faculty of Fisheries and Protection of Waters
Jihočeská univerzita v Českých Budějovicích
University of South Bohemia in České Budějovice

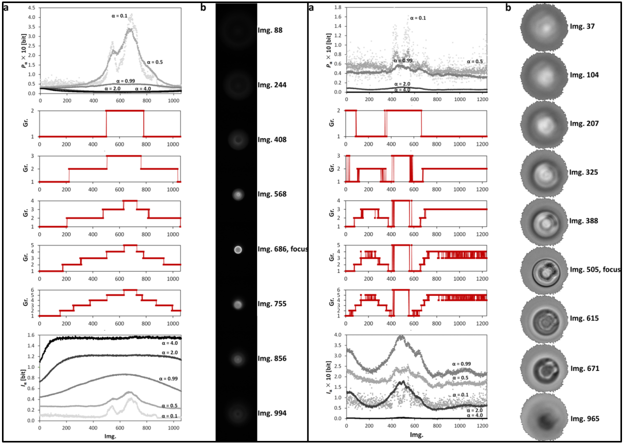


Figure 8: ***Left panel:*** *Clustering of a z-stack of grayscale microscopic images of a microring obtained using a fluorescence mode.*
Right panel: *Clustering of a z-stack of grayscale microscopic images in diffraction mode.*

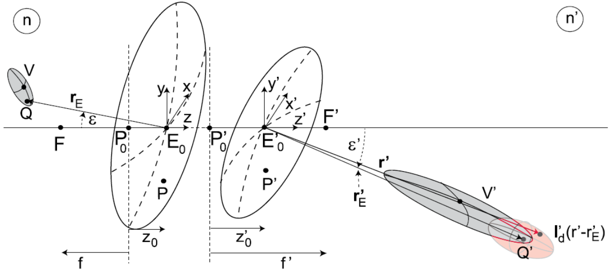


Figure 9: *In the real microscopic image of multiple objects the transfer function is represented by multiple self-simular responses from multiple different objects. This is the origin of multifractality in the image. The multifractality originating from all imaged objects is to a different extent present in all parts of the image. The* $\Omega_{\alpha}^{(l\to m)} = 0$ *may, in some positive cases, represent the location of the object and its extreme value (darkest or brightest voxel) is the centroid of the response of the electromagnetic field to the interaction with the objects, the* ***electromagnetic centroid****.*

Conclusions

PIG ($\Gamma_{\alpha,i}$), PIE (H_α), PIED (Ξ_α), PDG $\Omega_\alpha^{(l\to m)}$, PDGE I_α and PDGED P_α are useful tools for classification of multidimensional datasets.
They have potential in image enhancement, feature extraction, image sorting, data compression etc.
And there is a suspicion that the non-equilibrium thermodynamics may be more naturally established using them, but that would mean to introduce them to the equilibrium thermodynamics first.

Thank you for your attention!