Article

Fractional Order Generalized Information

José Tenreiro Machado

Institute of Engineering, Polytechnic of Porto, Department of Electrical Engineering, Rua Dr. António Bernardino de Almeida, 431, 4200-072 Porto, Portugal; E-Mail: jtm@isep.ipp.pt; Tel.: +351-22-8340500; Fax: +351-22-8321159

Received: 3 April 2014; in revised form: 16 April 2014 / Accepted: 21 April 2014 / Published: 24 April 2014

Abstract: This paper formulates a novel expression for entropy inspired in the properties of Fractional Calculus. The characteristics of the generalized fractional entropy are tested both in standard probability distributions and real world data series. The results reveal that tuning the fractional order allow an high sensitivity to the signal evolution, which is useful in describing the dynamics of complex systems. The concepts are also extended to relative distances and tested with several sets of data, confirming the goodness of the generalization.

Keywords: information; entropy; fractional calculus

1. Introduction

During the last decades the scientific community paid considerable attention to the generalization of concepts such as information, entropy [1–4] and differentiation [5–8]. Entropy was introduced in thermodynamics by Clausius and Boltzmann and was later adopted by Shannon and Jaynes in information theory [9–11]. Fractional Calculus (FC) was introduced by Leibniz in mathematics and found application in the areas of biology, physics and engineering [12–18]. The progress motivated the formulation of novel entropy indices and fractional operators, often relaxing some properties and allowing their application in complex dynamical systems [19–21].

The generalized concepts motivate further developments and new research avenues emerge. Bearing these ideas in mind, the present study combines both concepts and is organized as follows. Section 2 introduces entropy and fractional calculus in order to formulate the new generalized fractional entropy. Section 3 applies the new index in several types of data, namely two mathematical induced series, the digits of number π [22] and the Weierstrass function, two financial time series, the Dow Jones Industrial Average and the Europe Brent Spot Price [23,24], and one genomic series, the Human chromosome.
Y [25]. The results are analysed and distinct entropy formulations, for several fractional orders, compared. Section 4 expands the proposed index towards the concepts of distance. The Kullback-Leibler and Jensen-Shannon divergence measures are revisited and rewritten in the light of the fractional perspective. The performance of the index is tested with two sets of data, namely 13 irrational numbers and the whole 24 Human chromosomes, adopting the fractional order that reveals higher sensitivity. Finally, Section 5 outlines the main conclusions.

2. Fractional Generalization of Entropy

Information theory was developed by Claude Shannon in 1948 [26,27] and has been applied in many scientific areas. The fundamental cornerstone is the information content of some event having probability of occurrence \( p_i \):

\[
I(p_i) = - \ln p_i
\]

(1)

The expected value, called Shannon entropy [28,29], becomes:

\[
S = E(- \ln p) = \sum_i (- \ln p_i)p_i
\]

(2)

where \( E(\cdot) \) denotes the expected value operator.

Expression (2) obeys the four Khinchin axioms [30,31] and several generalizations of entropy have been proposed, obeying only a sub-set of them.

Recently Ubriaco brought together information theory and FC and proposed [32] the expression:

\[
S_q = E[(- \ln p)^q] = \sum_i (- \ln p_i)^q p_i
\]

(3)

where \( 0 \leq q \leq 1 \) denotes the “order” so that \( q = 1 \) yields Expression (2). This formulation obeys the same properties as the Shannon entropy except additivity and is the expected value of information content given by:

\[
I_q(p_i) = (- \ln p_i)^q
\]

(4)

It is well known in FC the adoption of a power function for obtaining intermediate values, that is, for “fractionating” classical integer operators. In brief, the Laplace transform of the fractional derivative of order \( \alpha \in \mathbb{R} \) of a signal \( x(t) \) with zero initial conditions is given by:

\[
\mathcal{L}\{0D^\alpha x(t)\} = s^\alpha \mathcal{L}\{x(t)\}
\]

(5)

where \( t \) represents time, and \( \mathcal{L}\{\cdot\} \) and \( s \) denote the Laplace operator and variable, respectively.

This property motivated the approximation of fractional derivatives by expanding the factor \( s^\alpha \) both with the Fourier and the \( \mathcal{Z} \) transforms [33,34]. However, the adoption by means of a power function is related with transforms and we can design a distinct fractional approach for information and entropy. In fact, we can think of Shannon information \( I(p_i) = - \ln p_i \) between the cases \( D^{-1}I(p_i) = p_i (1 - \ln p_i) \) and \( D^1I(p_i) = - \frac{1}{p_i} \), which, in the perspective of FC, leads to the proposal of information and entropy of order \( \alpha \in \mathbb{R} \) given by [35]:

\[
I_\alpha(p_i) = D^\alpha I(p_i) = - \frac{p_i^{-\alpha}}{\Gamma(\alpha + 1)} [\ln p_i + \psi(1) - \psi(1 - \alpha)]
\]

(6)
\[ S_\alpha = \sum_i \left\{ - \frac{p_i^{-\alpha}}{\Gamma(\alpha + 1)} \left[ \ln p_i + \psi(1) - \psi(1 - \alpha) \right] \right\} p_i \] (7)

where \( \Gamma(\cdot) \) and \( \psi(\cdot) \) represent the gamma and digamma functions.

Expression (7) fails to obey some of the Khinchin axioms with exception of the case \( \alpha = 0 \) that leads to the classical Shannon entropy. This behaviour is in line with what occurs in FC, where fractional derivatives fail to obey some of the properties of integer-order operators. By other words, in both cases, by generalizing operators we loose some classical properties.

Figure 1 shows the locus of \( I_q \) versus \( (q,p) \), \( 0 \leq q \leq 1 \), and \( I_\alpha \) versus \( (\alpha,p) \), \( -1 \leq \alpha \leq 1 \). We observe that \( I_q \) has a smaller amplitude excursion than \( I_\alpha \). Moreover, we verify that \( I_q \) takes not only positive, but also negative values for \( \alpha > 0 \). Therefore, Expression (6) assumes also the assumption that we can have negative information, that, for a given value of \( \alpha > 0 \), can be interpreted as derived from “misleading events”. While exploring the concept of “deception” is not the objective of the present paper, we should note that related ideas were addressed, in abstract terms, in the scope of negative probabilities [36–41] and, in practical terms, in the scope of robotics [42]. In short, we can say that the parameter \( \alpha \) allows us to tune the level of confidence of the information varying from positive (trustworthy) up to negative (deceptive) information.

**Figure 1.** Variation of information: \( I_q \) versus \( (q,p) \), \( 0 \leq q \leq 1 \) (left) and \( I_\alpha \) versus \( (\alpha,p) \), \( -1 \leq \alpha \leq 1 \) (right).

In order to illustrate the behaviour of the new index and to compare the two approaches Figure 2 depicts the entropies \( S_q \) and \( S_\alpha \) for the uniform, Poisson \( (\alpha = 2) \), geometric \( (p = 0.3) \), binomial \( (p = 0.3) \), and Benford probability distributions. We verify that \( S_q \) has a much smaller variation with \( q \) than \( S_\alpha \) with \( \alpha \). There is a large similarity between the shape of the curves for \( 0 < q \leq 1 \) and \( -0.5 < \alpha \leq 0 \). This is natural since \( S_q \) tends to the traditional entropy when \( q \to 1 \), while \( S_\alpha \) tends to the traditional entropy when \( \alpha \to 0 \). Furthermore, we verify that \( S_\alpha \) has maxima for \( 0.07 < \alpha < 0.23 \) and reaches null values for \( 0.62 < \alpha < 0.68 \). Therefore, in a practical application we can adopt values for \( \alpha \) in the first range if information is reliable, or we can consider values of \( \alpha \) in the second range if data contains misleading information.
Figure 2. Entropy variation for the uniform, Poisson ($\alpha = 2$), geometric ($p = 0.3$), binomial ($p = 0.3$), Benford probability distributions: $S_{q}^{bin}$ versus $q$ (left) and $S_{\alpha}^{bin}$ versus $\alpha$ (right).

Usually it is of interest to investigate the evolution for a binary distribution, that in our case leads to the expressions:

$$S_{q}^{bin} = p (- \ln p)^q + (1 - p) [ - \ln (1 - p)]^q$$

$$S_{\alpha}^{bin} = p \left\{ - \frac{p^\alpha}{\Gamma (\alpha + 1)} [ \ln (p) + \psi (1) - \psi (1 - \alpha)] \right\}$$

$$+ (1 - p) \left\{ - \frac{(1 - p)^\alpha}{\Gamma (\alpha + 1)} [ \ln (1 - p) + \psi (1) - \psi (1 - \alpha)] \right\}$$

Figure 3 shows the locus of $S_q$ and $S_{\alpha}$ versus $(p,q)$, $0 \leq q \leq 1$, and $(\alpha,p)$, $-1 \leq \alpha \leq 1$, respectively. In both cases we have a symmetrical variation relatively to $p = 0.5$, but $S_q$ is less sensitive than $S_{\alpha}$ to the variation of the order. In the case of $S_{\alpha}$ we observe that the chart passes from convex to concave in the region of $\alpha = 0.5$.

3. Application of the Generalized Entropy

This section applies $S_q$ and $S_{\alpha}$ to the mathematical constant $\pi$, the Weierstrass function, the Dow Jones Industrial Average (closing values) and the Europe Brent Spot Price (USD per barrel) financial time series, and one genomic series, the Human chromosome Y. The mathematical constant $\pi$ is expanded in base 10, and each digit is considered separately in the series. In the Weierstrass function, $f (\xi) = \sum_{n=0}^{\infty} a^n \cos (b^n \pi \xi)$ are adopted the parameters $a = 0.5$, $b = 3$ and the range $-2 \leq \xi \leq 2$. The two financial series correspond to daily values during the period 18 May 1987 up to 14 March 2014. In the four cases we adopt a total of $L = 7000$ data values. For the calculation of the histograms of relative frequency a non-overlapping sliding time window of $W = 100$ points is adopted, producing a total of $k = 1, \cdots, 70$ samples. In the case of the genomic series we have four bases denoted {A,C,T,G} that are sampled in groups of 3 producing histograms with $4^3$ bins. A small percentage of triplets involving the symbol N (considered as “not useful” in genomics) are not analysed. Therefore, a sequence of size $L = 872 \cdot 10^4$ is adopted and two distinct non-overlapping sliding windows, of
$W_1 = 10,000$ and $W_2 = 124,571$ points each, are considered producing a total of $k = 1, \cdots, 872$ and $k = 1, \cdots, 70$ samples, respectively.

**Figure 3.** Variation of entropy: $S^\text{bin}_q$ versus $(q, p)$, $0 \leq q \leq 1$ (left) and $S^\text{bin}_\alpha$ versus $(\alpha, p)$, $-1 \leq \alpha \leq 1$ (right).

Figures 4 and 5 represent $S_q$ versus $(q, p)$, and $S_\alpha$ versus $(\alpha, p)$, for the $\pi$ series and the Weierstrass function, respectively. We observe that $S_q$ has a low sensitivity to the dynamics of the series exhibiting significant variations only for $q$ close to one, that is, when it reduces to the Shannon entropy. On the other hand, $S_\alpha$ detects clearly dynamical variations, being particularly sensitive in the region $0 < \alpha < 0.6$.

Figures 6 and 7 depict the plots of $S_q$ and $S_\alpha$ for the Dow Jones Industrial Average and Europe Brent Spot Price, respectively. We verify a behaviour similar to the one pointed out previously.

**Figure 4.** Entropy variation for the $\pi$ series: $S_q$ versus $(q, p)$, $0 \leq q \leq 1$ (left) and $S_\alpha$ versus $(\alpha, p)$, $-1 \leq \alpha \leq 1$ (right).
Finally, Figures 8 and 9 show $S_q$ and $S_\alpha$ for the Human chromosome Y, with the only difference being the size and number of sliding windows. As previously the higher sensitivities occur for $q = 1$ with $S_q$ and for $\alpha = 0.5$ with $S_\alpha$. The sliding window $W_1$ is more appropriate for highlighting dynamical evolutions than window $W_2$ that is considerable large and leads to an “averaging” of the information content of the chromosome series.

It is interesting to note that the average entropy over the complete data series characterizes the type of embedded information. In fact, the maxima values are $(\alpha, S_\alpha^{av}) = (0.225, 2.52)$, $(\alpha, S_\alpha^{av}) = (0.375, 4.08)$ and $(\alpha, S_\alpha^{av}) = (0.40, 0.217)$, for the $\pi$, {Weierstrass, Dow Jones Industrial Average, Europe Brent Spot Price} and {Human chromosome Y} (both windows) data series, respectively. The results remain identical for other numerical constants and chromosomes to be discussed in the next section.
Figure 7. Entropy variation for the Europe Brent Spot Price time series: $S_q$ versus $(q,p)$, $0 \leq q \leq 1 \text{ (left)}$ and $S_\alpha$ versus $(\alpha,p)$, $-1 \leq \alpha \leq 1 \text{ (right)}$.

Figure 8. Entropy variation for the Human chromosome Y, $W_1 = 10,000$: $S_q$ versus $(q,p)$, $0 \leq q \leq 1 \text{ (left)}$ and $S_\alpha$ versus $(\alpha,p)$, $-1 \leq \alpha \leq 1 \text{ (right)}$.

Figure 9. Entropy variation for the Human chromosome Y, $W_2 = 124,571$: $S_q$ versus $(q,p)$, $0 \leq q \leq 1 \text{ (left)}$ and $S_\alpha$ versus $(\alpha,p)$, $-1 \leq \alpha \leq 1 \text{ (right)}$. 
4. Application of the Generalized Entropy

In this section we explore further the concept of generalized fractional information and entropy. We start by recalling the Kullback-Leibler divergence of $Q$ from $P$ defined as [43–47]:

$$D_{KL}(P \parallel Q) = \sum_i p_i \ln \frac{p_i}{q_i}$$  \hspace{1cm} (10)

The Jensen-Shannon divergence $JSD(P \parallel Q)$ is defined as:

$$JSD(P \parallel Q) = \frac{1}{2} [D_{KL}(P \parallel M) + D_{KL}(P \parallel M)]$$  \hspace{1cm} (11)

where $M = \frac{P+Q}{2}$.

Alternatively, we can calculate $JSD(P \parallel Q)$ as:

$$JSD(P \parallel Q) = \frac{1}{2} \left[ \sum_i p_i \ln p_i + \sum_i q_i \ln q_i \right] - \sum_i m_i \ln m_i$$  \hspace{1cm} (12)

Having in mind Expressions (4), (6) and (12), the fractional $JSD$ can be written as:

$$JSD_q(P \parallel Q) = \frac{1}{2} \sum_i p_i (- \ln p_i)^q + \frac{1}{2} \sum_i q_i (- \ln q_i)^q - \sum_i m_i (- \ln m_i)^q$$  \hspace{1cm} (13)

$$JSD_\alpha(P \parallel Q) = \frac{1}{2} \sum_i p_i \left\{ - \frac{p_i^{-\alpha}}{\Gamma (\alpha + 1)} [\ln p_i + \psi (1) - \psi (1 - \alpha)] \right\} +$$

$$\frac{1}{2} \sum_i q_i \left\{ - \frac{q_i^{-\alpha}}{\Gamma (\alpha + 1)} [\ln q_i + \psi (1) - \psi (1 - \alpha)] \right\}$$

$$- \sum_i m_i \left\{ - \frac{m_i^{-\alpha}}{\Gamma (\alpha + 1)} [\ln m_i + \psi (1) - \psi (1 - \alpha)] \right\}$$  \hspace{1cm} (14)

In order to illustrate the fractional-order distance we consider two examples, namely the set $A$ of $n = 13$ irrational numbers and the set $B$ of $n = 24$ Human chromosomes. Set $A$ consists of the numbers $P_i$ ($\pi = 3.141 \cdots$), Nepper ($e = 2.718 \cdots$), Euler-Mascheroni ($\gamma = 0.577 \cdots$), Catalan ($G = 0.915 \cdots$), Hilbert or Gelfond-Schneider ($2^{\sqrt{2}} = 2.665 \cdots$), Khinchin ($K_0 = 2.685 \cdots$), Golden ratio ($\varphi = \frac{1 + \sqrt{5}}{2} = 1.618 \cdots$), ln 2, ln 3, ln 5, $\sqrt{2}$, $\sqrt{3}$ and $\sqrt{5}$ labelled in the sequel as $\{P, N, Eul, Cat, Hil, Gol, Ln2, Ln3, Ln5, St2, St3, St5\}$. Set $B$ consists of the whole set of Human chromosomes labelled in the sequel as $\{Hu1, ..., Hu22, HuX, HuY\}$. The irrational numbers are expanded up to 7000 digits and, for each one, groups of two digits feed $10^2$ bins of histograms of relative frequency of occurrence. On the other hand, the chromosome bases are read in triplets feeding $4^3$ bins of histograms of relative frequency of occurrence. In both cases, a comparison $n \times n$ symmetrical matrix $D$ of element to element relative distances is constructed, adopting the indices $JSD_\alpha$ and $JSD_\alpha$. For simplifying comparisons all distances were converted to the interval between zero (minimum distance) and one (maximum distance). The results are visualized by means of Phylip [48,49] (plots using options “neighbor” and “drawtree”), a package of programs for inferring phylogenies. These algorithms produces a tree based on matrix $D$, trying to accommodate the distances into the two dimensional space.
Figures 10 and 11 show the trees for sets $\mathcal{A}$ and $\mathcal{B}$ based of distances (13) and (14). We verify that not only the charts are qualitatively of the same type, but also that the generalization leads to results compatible with those produced by distinct methods [50–52] which confirms the goodness of the proposed concept.

**Figure 10.** Tree (Phylip with algorithm “neighbor” and visualization by “drawtree”) of the set $\mathcal{A}$ of 13 irrational numbers, compared by means of the indices $I_q$, $q = 1$ (left) and $I_\alpha$, $\alpha = 0.5$ (right).

**Figure 11.** Tree (Phylip with algorithm “neighbor” and visualization by “drawtree”) of the set $\mathcal{B}$ of 24 Human chromosomes compared by means of the indices $I_q$, $q = 1$ (left) and $I_\alpha$, $\alpha = 0.5$ (right).

5. Conclusions

This paper presented a generalization of the concept of entropy inspired in the properties of Fractional Calculus. The novel index follows the recent trend in expanding the scope of application of both mathematical tools, by relaxing some properties and allowing their application in new scientific areas. The generalized fractional entropy was first adopted with several typical probability distributions. In a second phase the index was also applied to several types of data, namely of mathematical, financial and biological nature. It was verified that the proposed entropy leads to a higher sensitivity to the signal evolution being useful in describing the dynamics of complex systems. Furthermore, the proposed generalization embeds the concept of positive and negative information, that is, with data either reliable or misleading, allowing the extension of entropy for deceptive cases. The new formulation is then
extended for measuring relative distances and tested with two distinct sets of data. The results reveal the goodness of the generalized fractional information concept.

Acknowledgments

The author thanks the anonymous reviewers for their constructive comments.

Conflicts of Interest

The author declares no conflict of interest.

References


