Entropy and Statitical Complexity Measure

Osvaldo

October 18, 2021

Abstract

This must be a new Section 2, then all the resting section in the text must be increase in one unity. The main idea in this Section 2, is describe how we evaluate the normalized Shannon entropy and how we define the statistical complexity, and give some of the properties that we consider in order to choose it.

1 Introduction

No modifications by the moment.

2 The Information Theory descriptors

According to Information Theory, entropy is a relevant measure of order and disorder for many systems, including dynamical ones (Cover and Thomas, 1991). Indeed, Kolmogorov (1958) and Sinai (1959) converted the Shannon's Information Theory (1948) into a powerful tool for the study of dynamical systems.

Let a discrete real time series $\mathcal{X} = \{x_j, j = 1, \ldots, M\}$ under analysis. There are many possibilities to assign a PDF to given time series. In particular, we given one in consideration in the next section. In the meantime suppose that the PDF is discrete and given by $P = \{p_j, j = 1, \ldots, N\}$ with $p_j \in \mathbb{R}$ and $\sum_{j=1}^{N} p_j = 1$. We define various Information Theory descriptors as follow:

(i) Normalized Shannon Entropy, $H[P]$:
The Shannon entropy is widely used as first natural approach to quantify the information content of a system. Let $S[P]$ the Shannon entropy (Shannon, 1948) given by

$$S[P] = -\sum_{j=1}^{N} p_j \ln(p_j) \ . \tag{1}$$

Its value is regarded as the measure of uncertainty associated to the physical process described by P. IF $S[P] = 0$, our knowledge of the underlying process described by the PDF is maximal. In contrast, our knowledge is minimal for a uniform distribution. The maximum value $S_{max} = \ln(N)$ is obtained for uniform probability distribution, $P_e = \{p_j = 1/N, \forall j = 1, \ldots, N\}$, where N, total number of states of the system in phase space.

The "Normalized Shannon Entropy" ($0 \le H[P] \le 1$) can be defined as

$$H[P] = S[P]/S_{max} \ . \tag{2}$$

We would like to remark that entropy measures do not quantify *the degree of structure or patterns* present in the process and measures of statistical or structural complexity are necessary to capture properties related to organization.

(ii) Statistical Complexity, $C[P]$:
We start from the López-Ruiz, Mancini and Calbet (1995) measure of statistical complexity defined as

$$C^{(LMC)}[P] = H[P] \cdot Q_E[P, P_e] \ , \tag{3}$$

in which H represent the normalized Shannon entropy and Q_E is the "disequilibrium-distance" between the probability distribution of the system P and the equilibrium uniform distribution P_e, taken as Euclidean norm.

It has been pointed out by Crutchfield and co-workers (1998) that the LMC complexity measure is marred by some troublesome characteristics:

(1) it is neither an intensive nor an extensive quantity.

(2) it vanishes exponentially in the thermodynamic limit for all one-dimensional finite-range systems.

Also, the above authors forcefully argue that a reasonable complexity measure should

(3) be able to distinguish among different degrees of periodicity.

(4) vanishes only for unity periodicity.

The product functional form for the complexity measure makes it impossible to overcame the second deficiency mentioned above. In previous works Rosso and co-workers shown that, after performing some suitable changes in the definition of the disequilibrium-distance, by means of utilization of either Wootters distance (Martín et al. 2003) or Jensen's divergence (Lamberti et al. 2004), one is in position to obtain a complexity measure that is:

(i) able to grasp essential details of the dynamics (i.e., chaos, intermitency, etc).

(ii) capable of discerning between different degrees of periodicity, and

(iii) an intensive quantity if Jensen's divergence is used.

Then the present statistical complexity measure is defined as

$$C[P] = H[P] \cdot Q_J[P, P_e] \ , \tag{4}$$

where $H[P]$ is the normalized Shannon entropy and $Q_J[P, P_e]$ is the above referred "disequilibrium" defined in terms of the extensive Jensen-Shannon divergence (Lamberti et al. 2004)

$$Q_J[P, P_e] = Q_0 \cdot \{S[(P + P_e)/2] - S[P]/2 - S[P_e]/2\} \ , \tag{5}$$

with Q_0 a normalization constants ($0 \le Q_J \le 1$) that reads

$$Q_0 = -2 \left\{ \left(\frac{N+1}{N} \right) \ln(N+1) - 2\ln(2N) + \ln N \right\}^{-1} \ . \tag{6}$$

We see that the disequilibrium Q_J is an intensive quantity that reflects on the system's "architecture¨, being different from zero only if there exist "privileged", or "more likely" states among the accessible ones. $C[P]$ quantifies the presence of correlational structures as well.

The opposite extremes of perfect order and maximal randomness possess no structure to speak of and, as consequence, $C[P] = 0$. In between these two special instances a wide range of possible degree of physical structure exist, degrees that should be reflected in the features of the underlying probability distribution. In the case of ideal "radom number generator" the "ideal" values are "$H[P] = 1$" and "$C[P] = 0$¨.

In addition, we see that the Statistical Complexity Measure is also a normalized quantity, that is $0 \leq C[P] \leq 1$, and has the intensive property found in many thermodynamic quantities. We stress that fact that the statistical complexity measure defined above is the product of two normalized entropies (the Shannon entropy and Jensen-Shannon divergence), but it is a nontrivial function of the entropy because it depends of two different probabilities distributions, i.e., the one corresponding to the state of the system, P, and the uniform distribution, P_e. In particular, Rosso and co-workers (2006), shown that for each given value of normalized Shannon entropy, the corresponding values of statistical complexity, varies between C_{min} and C_{max} values.

(iii) The Entropy-Complexity Plane, $H \times C$:

In statistical mechanics one is often interested in isolated systems characterized by an initial, arbitrary, and discrete probability distribution. Evaluation towards equilibrium is to be described, as the overriding goal. At equilibrium, we can think, without loss of generality, that this state is given by the uniform distribution P_e. In order to study the time evolution of the statistical complexity measure a diagram of C versus time t can be used. However, as we know, the second law of thermodynamic states that in isolated system entropy grows monotonically with time ($dH/dt \geq 0$) (Plastino and Plastino, 1996). This implies that H can be regarded as arrow of time, so that an equivalent way to study the time evolution of complexity is to plot C versus H. In this way, the normalized entropy-axis substitutes the time-axis. This kind of diagram $H \times C$ has been used also to study changes in the dynamics of system originated by modifications of some characteristics parameters (Referencias varias),(Feldman et al., 2004).

References

Cover, T. M.; Thomas, J. A. Elements of Information Theory, Wiley, New York, 1991.

Kolmogorov, A. N. A new metric invariant of transitive dynamic system and automorphysms in Lebesgue spaces. Dokl. Akad. Nauk. SSSR 119 (1958) 861–864.

Sinai, Y. G. On the concept of entropy of dynamical system Dokl. Akad. Nauk. SSSR 124 (1959) 768–771.

Shannon, C. E. A mathematical theory of communication. Bell Syst. Technol J. 27 (1948) 379–423 and 623–656.

Feldman, D. P.; Crutchfield, J. P. Measures of statistical complexity, Why ? Phys. Lett. A 238 (1998) 244–252.

Lamberti, P. W.; Martín, M. T.; Plastino, A.; Rosso, O. A. Intensive entropic non-triviality measure. Physica A 334 (2004) 119–131.

Martín, M. T.; Plastino, A.; Rosso, O. A. Statistical complexity and disequilibrium. Phys. Lett. A 331 (2003) 126–132.

De Micco, L.; González, C. M.; Larrondo, H. A,; Martín, M. T.; Plastino, A.; Rosso, O. A. Randomizing nonlinear maps via symbolic dynamics Physica A 387 (2008) 3373–3383.

De Micco, L.; Larrondo, H. A,; Plastino, A.; Rosso, O. A. Quantifiers for randomnness of pseudo-random number generators. Phil. Trans. R. Soc. A 367 (2009) 3281–3296.

Feldman, D. P.; Tague, C. S.; Crutchfield, J. P. The organization of intrinsic computation: Complexity-entropy diagrams and diversity of natural information processing. Chaos 18 (2008) 043106-1, 043106-15.

Plastino, A. R.; Plastino, A. Symmetries of the Focker-Plack equation and the Fisher-Frieden arrow of time. Phys. Rev. E 54 (1996) 4423–4326.