Extension of Gatlin's Informational Divergence to Markovian Stochastic Processes

GUY JUMARIE

Department of Mathematics, University of Quebec at Montreal, Montreal, Canada

Accepted in final form 10 March 1997

Abstract. In the information theoretic framework, to some extent, the complexity of a system can be measured by its informational divergence which is the difference between the maximum possible value of its entropy and the actual value of the latter. In her analysis of the DNA chain, Gatlin put in evidence two different divergences. One which drops the mutual dependences between the bases (A, C, G, T) and the other one which explicitly refers to this dependence via conditional entropies, and thus, provides a measure of the structural complexity of the system. One shows how the explicit form of this structural divergence can be obtained for discrete and continuous Markovian stochastic processes, and in the latter case, as expected, this divergence is invariant under a transformation of variables.

Key words: Complexity, Informational divergence, Redundancy, Path entropy, Stochastic processes

1. Introduction

Defining the complexity of a system is not a simple task [1, 2]. Many models have been suggested, and all of them presuppose that complexity is neither complete order nor complete disorder [3–6]. Also it has been conjectured that complexity tends to appear close to an instability point [7].

Here, we shall consider the complexity of a system S as related to the amount of information stored in it: The more the system contains information, the more it exhibits complexity. But information here is exactly transinformation in the Shannon sense. Clearly, it is the difference $I(S) = H_m(S) - H(S)$ between the maximum value $H_m(S)$ of the (Shannon) entropy of the system and the actual value $H(S)$ of the latter. $I(S)$ can be thought of as the information due to the organization of the system. When the system is completely disorganized, its entropy is $H_m(S)$; and this entropy decreases to $H(X)$ as soon as it exhibits some features of organization. $I(S)$ has been already used by some authors [see, for instance, refs. 8–10]; and here we shall show it can apply to stochastic processes.

The paper is organized as follows. In the next section, we shall give a brief description of Gatlin's papers [9, 10] in which she proposed two kinds of informational divergences: one which is related to the complexion of the system, irrespective of its internal structure; and another one which explicitly refers to this internal structure via conditional probabilities. It is this second divergence which

we extend to stochastic processes. We first obtain an expression for the path entropy of Markovian stochastic processes, and then use it to define the divergence, and we show that it is invariant under transformation of variables.

2. Background on Some Results

2.1. PRELIMINARY DEFINITIONS

In her papers [9, 10], Gatlin focused on the amount of *information stored* in DNA's chains. In order to apply (Shannon) information theory, it is necessary to randomize the problem, and to this end, one assumes that the four bases A (adenine), C (cytosine), G (guanine) and T (thynine) are the different states of a given random system S .

- (i) If one defines (the symbol $:=$ means that the left side is defined by the right one)

$$p_1 := \text{pr}\{A\}, \quad p_2 := \text{pr}\{C\}, \quad p_3 := \text{pr}\{G\}, \quad p_4 := \text{pr}\{T\}, \quad (2.1)$$

where p_i ($i = 1, \dots, 4$) is the frequency which is experimentally observed, one can define the entropy

$$H_{\text{obs}}(S) := - \sum_{i=1}^4 p_i \ln p_i. \quad (2.2)$$

- (ii) Next, if one considers a sequence $s_1, s_2, s_3, \dots, s_n$ of bases in a chain (of which, for instance, a special case could be AATCCCGT...), its Shannon entropy is

$$\begin{aligned} H_n(S) &= H(s_1, s_2, \dots, s_n) \\ &= H(s_1) + H(s_2|s_1) + \dots + H(s_n|s_1, s_2, \dots, s_{n-1}). \end{aligned} \quad (2.3)$$

Therefore, one can define the average value

$$H_R(s) := \lim_{n \rightarrow \infty} \frac{1}{n} H_n(S) \quad \text{as} \quad n \uparrow \infty, \quad (2.4)$$

where the subscript R emphasizes that the corresponding entropy so obtained takes account of the redundancy in the DNA chain.

2.2. INFORMATION STORED IN DNA CHAINS

Gatlin defined the amount of information stored in the DNA chain by the expression

$$I = D_1 + D_2, \quad (2.5)$$

where D_1 and D_2 denote two informational divergences defined as follows:

(i) D_1 is equal to

$$D_1 := H_{\max}(S) - H_{\text{obs}}(S), \quad (2.6)$$

where $H_{\text{obs}}(S)$ is given by the Equation (2.2), and where $H_{\max}(S) = \ln 4$ is the maximum possible value of $H_{\text{obs}}(S)$ which is achieved when $p_1 = p_2 = p_3 = p_4 = \frac{1}{4}$.

(ii) Regarding D_2 , it is defined by the expression

$$D_2 := H_{\text{obs}}(S) - H_R(S), \quad (2.7)$$

which provides

$$I = H_{\max}(S) - H_R(S). \quad (2.8)$$

If one refers to the redundancy as defined by Shannon, namely

$$R := 1 - [H_R(S)/H_{\max}(S)], \quad (2.9)$$

one finds that

$$I = RH_{\max}(S). \quad (2.10)$$

2.3. DIVERGENCE AND STUCTURAL COMPLEXITY

As pointed out in the introduction, D_1 , D_2 and $D_1 + D_2$ can be thought of as measures of different type of complexities for the DNA chain. D_1 measures the complexity defined by the observed system, or observed entropy $H_{\text{obs}}(S)$, while D_2 measures the complexity due to the redundancy in the DNA chain, which is defined by the conditional probabilities involved in the conditional entropies of the Equation (2.3). In this way, D_2 could be considered as a *structural divergence* characterizing the *structural complexity* of the system. Our purpose, in the following, is to show how one can extend this structural divergence to Markovian stochastic processes.

3. Path Entropy of Stochastic Processes

The system S under consideration is described by the one-dimensional stochastic difference equation.

$$X_{n+1} = f(X_n, t_n) + g(X_n, t_n)w_n, \quad (3.1)$$

where X_0 has a given fixed value, $t_n = nT$ with T constant, and w_n is a Gaussian white noise with zero mean and the standard deviation σ_n . The following lemma is straightforward.

LEMMA 3.1. *The trajectory (or path) $(X_1, X_2, \dots, X_N)$ generated by the Equation (3.1) is a N -random vector of which the (Shannon) entropy is*

$$H[X(\cdot); 1, N] = \sum_{k=1}^N \int_R p(x_k, t_k) \ln[2\pi e \sigma_k^2 g^2(x_k, t_k)]^{1/2} dx_k, \quad (3.2)$$

where $p(x_k, t_k)$ is the probability density of X_k .

Proof. (i) Due to the Markovian nature of the process, one has the equality (H refers to the Shannon entropy of random variables)

$$H(X_0, X_1, \dots, X_N) = H(X_0) + \sum_{k=1}^N H(X_k | X_{k-1}). \quad (3.3)$$

(ii) $(X_{k+1} | X_k)$ is a Gaussian random variable with zero mean and the variance is $\sigma_k^2 g^2(x_k, t_k)$. Its entropy, for a given x_k , is thus $\ln[2\pi e \sigma_k^2 g^2(x_k, t_k)]^{1/2}$. Therefore the conditional entropy is

$$H(X_{k+1} | X_k) = \int_R p(x_k, t_k) \ln[2\pi e \sigma_k^2 g^2(x_k, t_k)]^{1/2} dx_k. \quad (3.4)$$

(iii) Since there is no uncertainty about the initial value X_0 , we shall consider only the contribution

$$H = \sum_{k=1}^N H(X_k | X_{k-1}), \quad (3.5)$$

as the result, via Equation (3.4). □

4. Structural Divergence of Discrete Stochastic Processes

LEMMA 4.1. *Consider the discrete stochastic process X_k , in Equation (3.1), of which the state probability density is $p(x_k, t_k)$. Let $I[X(\cdot); 1, N]$ denote the structural divergence on the range $[1, N]$, which is equal to the difference between the actual trajectory entropy $H[X(\cdot); 1, N]$ and the path entropy which would be obtained if all the X_k 's were mutually independent. Then, the divergence is defined by the expression*

$$\begin{aligned} I[X(\cdot); 1, N] \\ = - \sum_{k=1}^N \int_R p(x_k, t_k) \ln[(p(x_k, t_k)) [2\pi e \sigma_k^2 g^2(x_k, t_k)]^{1/2}] dx_k. \end{aligned} \quad (4.1)$$

Proof. (i) If all the X_k 's were mutually independent, we would have the equality

$$\begin{aligned} H(X_1, X_2, \dots, X_N) &= \sum_{k=1}^N H(X_k, t_k) \\ &= H_{\text{obs}}[X(\cdot); 1, N], \end{aligned} \quad (4.2)$$

with

$$H(X_k, t_k) = - \int_R p(x_k, t_k) \ln p(x_k, t_k) dx_k. \quad (4.3)$$

(ii) We then can calculate explicitly the difference

$$I[X(\cdot); 1, N] = H_{\text{obs}}[X(\cdot); 1, N] - H[X(\cdot); 1, N] \quad (4.4)$$

to obtain the result.

$I[X(\cdot); 1, N]$ is positive or zero, and it is equal to zero if and only if all the X_k 's are mutually independent. This property suggests the structural complexity of the process can be obtained from the redundancy

$$R_{\text{obs}}[X(\cdot); 1, N] = 1 - \frac{H[X(\cdot); 1, N]}{H_{\text{obs}}[X(\cdot); 1, N]}, \quad (4.5)$$

which yields

$$R_{\text{obs}}[X(\cdot); 1, N] = \frac{I[X(\cdot); 1, N]}{H_{\text{obs}}[X(\cdot); 1, N]}. \quad (4.6)$$

Illustrative example. Consider the stochastic differential equation

$$x_{n+1} = \alpha x_n + \beta^{1/2} w_n, \quad \beta > 0, \quad (4.7)$$

with $E\{w_n\} = 0$ and $\text{Var}\{w_n\} = 1$. A simple calculation yields

$$H[X(\cdot); 1, N] = N \ln(2\pi e\beta)^{1/2}, \quad (4.8)$$

$$\begin{aligned} H_{\text{obs}}[X(\cdot); 1, N] &= \sum_{k=1}^N \ln \left(2\pi e\beta \frac{1 - \alpha^{2k+2}}{1 - \alpha^2} \right)^{1/2} \\ &= H[X(\cdot); 1, N] + \sum_{k=1}^N \ln \left(\frac{1 - \alpha^{2k+2}}{1 - \alpha^2} \right)^{1/2}, \end{aligned} \quad (4.9)$$

and

$$I[X(\cdot); 1, N] = \sum_{k=1}^N \ln \left(\frac{1 - \alpha^{2k+2}}{1 - \alpha^2} \right)^{1/2}. \quad (4.10)$$

The complexity of the trajectory at time N is then measured by the redundancy

$$R_{\text{obs}}[X(\cdot); 1, N] = \left[1 + \frac{\sum_{k=1}^N \ln(2\pi e\beta)^{1/2}}{\sum_{k=1}^N \ln \left(\frac{1 - \alpha^{2k+2}}{1 - \alpha^2} \right)^{1/2}} \right]^{-1}. \quad (4.11)$$

If $\alpha = 0$, then as expected, $R_{\text{obs}}[X(\cdot); 1, N] = 0$, since then X_n is the Gaussian white noise itself. If $\beta = 0$, the redundancy is again equal to zero. Clearly, as long as we work in the randomness framework, deterministic dynamics do not exhibit complexity. In other words, the complexity defined by the structural divergence, measures the complexity due only to randomness. $\square$

5. Extension to Continuous Stochastic Processes

5.1. MAIN EQUATIONS

It is tempting to extend the preceding calculations to the continuous stochastic process defined by the stochastic differential equation

$$\dot{x} = f(X, t) + g(X, t)w(t), \quad (5.1)$$

but care must be exercised. Indeed, it is well known that the entropy of continuous variables is not the limiting form of entropy of discrete variables, i.e. $-\sum_i p_i \ln p_i$, and to circumvent this difficulty, one takes the limit of

$$H = - \sum_{k=1}^N p_k \ln(p_k / \mu_k), \quad (5.2)$$

where μ_k is a probability (referred to as an invariant measure) function proportional to the limiting density of the discrete point [see, for instance, ref. 11].

It is likely that we shall have similar trouble with the discretization in time here. So, at first glance, it seems wise to consider the mean value in time, to obtain

$$H_{\text{obs}}[X(\cdot); 0, T] = -\frac{1}{T} \int_0^T \int_R p(x, t) \ln p(x, t) \, dx \, dt, \quad (5.3)$$

$$H[X(\cdot); 0, T] = \frac{1}{T} \int_0^T \int_R p(x, t) \ln[2\pi e\sigma^2(t)g^2(x, t)]^{1/2} dx dt \quad (5.4)$$

and

$$\begin{aligned} I[X(\cdot); 0, T] \\ = -\frac{1}{T} \int_0^T \int_R p(x, t) \ln(p(x, t)[2\pi e\sigma^2(t)g^2(x, t)]^{1/2}) dx dt, \end{aligned} \quad (5.5)$$

where $p(x, t)$ is given by the Fokker–Planck–Kolmogorov equation

$$\partial_t p(x, t) = -\partial_x [f(x, t)p(x, t)] + \frac{1}{2}\partial_{xx} [g^2(x, t)\sigma^2(t)p(x, t)]. \quad (5.6)$$

5.2. INVARIANCE OF THE STRUCTURAL DIVERGENCE

The following result is obtained.

LEMMA 5.1. *Assume a change of variables $Y = \phi(X)$, where $\phi(\cdot)$ is a continuously differentiable function and X is the solution of Ito's stochastic differential Equation (5.1). Then one has the equality*

$$I[X(\cdot); 0, T] = I[Y(\cdot); 0, T], \quad (5.7)$$

for any positive T .

Proof. (i) The proof is a direct consequence of Ito's Lemma which states the equality

$$\dot{y} = (\phi_x f + \frac{1}{2}g^2\sigma^2\phi_{xx}) + \phi_x gw. \quad (5.8)$$

(ii) As a consequence of this equation, one then has

$$I[Y(\cdot); 0, T] = -\frac{1}{T} \int_0^T \int_R q(y, t) \ln(q(y, t)[2\pi e\sigma^2 g^2 \phi_x^2]^{1/2}) dx dt, \quad (5.9)$$

where $q(x, t)$ is the probability density of $Y(t)$.

(iii) We now remark that the equality

$$p(x, t) = q(\phi(x), t)|\phi'_x(x)| \quad (5.10)$$

obtains the result.

Remarks. $I[X(\cdot); 0, T]$ is invariant under transformation of variables, but this is not the case for the complexity R . In fact, the following equality holds:

$$H_{\text{obs}}[Y(\cdot); 0, T] = H_{\text{obs}}[X(\cdot); 0, T] + \int_0^T \int_R p(x, t) \ln |\phi'_x(x)| \, dx. \quad (5.11)$$

□

6. Outline of Applications

In many practical problems, the stochastic differential equation

$$\dot{x} = f(x, t) + g(x, t)w(t), \quad x(0) = x_0, \quad (6.1)$$

is obtained as the perturbation equation of the dynamics

$$\dot{x} = f(x, t), \quad x(0) = x_0. \quad (6.2)$$

In such a case the linearized dynamics (y holds for δx) is considered:

$$\dot{y}(t) = f'(x_r, t) + g(x_r, t)w(t), \quad (6.3)$$

where $x_r(t)$ is the reference dynamics defined by Equation (6.2).

Illustrative example. In order to have an idea of the kind of result obtainable, assume that the linearized equation is

$$\dot{y} = -\alpha(t)y + \beta(t)w(t), \quad (6.4)$$

with

$$E\{w\} = 0, \quad \text{Var}\{w(t)\} = 1.$$

(i) The corresponding Fokker–Planck–Kolmogorov equation is

$$\partial_t p(y, t) = \partial_y(\alpha y p) + \frac{1}{2} \partial_{yy}(\beta^2 p) \quad (6.5)$$

and in this special case, it is well known that $p(y, t)$ is the normal density of which the explicit expression is

$$p(y, t) = [2\pi v(t)]^{1/2} \exp \left\{ -\frac{[x - m(t)]^2}{2v(t)} \right\}, \quad (6.6)$$

where $m(t)$ and $v(t)$ are defined by the differential equations

$$\dot{m}(t) = -\alpha(t)m(t), \quad m(0) = 0, \quad (6.7)$$

$$\dot{v}(t) = -2\alpha(t)v(t) + \beta^2(t), \quad v(0) = 0. \quad (6.8)$$

(ii) $H_{\text{obs}}(X(\cdot); 0, T)$ is the entropy of the normal law, and is thus equal to

$$H_{\text{obs}}[X(\cdot); 0, T] = \frac{1}{T} \int_0^T \ln[2\pi e v(t)]^{1/2} dt. \quad (6.9)$$

(iii) According to Equation (5.4), the path entropy is

$$H[X(\cdot); 0, T] = \frac{1}{T} \int_0^T \ln[2\pi e \beta^2(t)]^{1/2} dt, \quad (6.10)$$

therefore the structural divergence is

$$I[X(\cdot); 0, T] = \frac{1}{T} \int_0^T \ln[v(t)/\beta^2(t)] dt. \quad (6.11)$$

7. Concluding Remarks

In the above derivation, for ease of exposition (a student would say for pedagogical reasons) we restricted ourselves to one-dimensional systems, but generalization to n - D systems is straightforward. The basic formulae to this end are the following.

(i) With standard notations, the entropy of the n - D normal density

$$p(x, t) = (2\pi)^{-(n/2)} \left| \Sigma \right|^{-(1/2)} \exp \left\{ -\frac{1}{2} (x - \mu)^T \Sigma^{-1} (x - \mu) \right\} \quad (7.1)$$

is

$$H(X, t) = \ln \left[\left| \Sigma \right|^{1/2} (2\pi e)^{n/2} \right]. \quad (7.2)$$

(ii) Consider the n - D stochastic differential equation

$$\dot{x}(t) = f(x, t) + G(x, t)w(t), \quad (7.3)$$

where x and f are n -vectors, G is $n \times n$, and $w(t)$ is an n -vector Gaussian white noise with zero mean and the covariance matrix $E\{w(t)w^T(t)\} = \Sigma(\tau)$. Then, its Fokker–Planck–Kolmogorov equation is

$$\frac{\partial p}{\partial t} = - \sum_{i=1}^n \frac{\partial (p f_i)}{\partial x_i} + \frac{1}{2} \sum_{i,j=1}^n \frac{\partial^2 [p (G \Sigma G^T)_{ij}]}{\partial x_i \partial x_j}. \quad (7.4)$$

(iii) As a result, the parallel of Equations (5.3) and (5.4) reads

$$H_{\text{obs}}[X(\cdot); 0, T] = -\frac{1}{T} \int_0^T \int_{R^n} p(x, t) \ln p(x, t) dx dt \quad (7.5)$$

and

$$H[X(.); 0, T] = \frac{1}{T} \int_0^T \int_{R^n} p(x, t) \ln \left[(2\pi e)^{n/2} \left| G \sum G^T \right|^{1/2} \right] dx dt. \quad (7.6)$$

Acknowledgments

This paper was written, while the author, who is a French citizen from the West Indies (Guadeloupe Island), resided in Canada.

References

1. Grassberger, P.: Toward a quantitative theory of self-generated complexity, *Int. J. Theoret. Phys.* **25** (1986), 907–938.
2. Haken, H.: *Information and self-organization*. Berlin: Springer Verlag, 1989.
3. Crutchfield, J.P. and Young K.: Computation at the onset of chaos. In: *Complexity, entropy and the physics of information*. New York: Addison-Wesley, 1990, pp. 32–55.
4. Huberman, B.A. and Hogg, T.: Complexity and adaptation, *Physica D.* **22** (1986), 376–384.
5. Langton, G.: Computation at the edge of chaos, *Physica D.* **22** (1990), 12–32.
6. Nicolis, G. and Prigogine, I.: *Exploring complexity*. New York: Freeman, 1989.
7. Sole, R.V., Manrubia, S.C., Luque, B., Delgado, J. and Bascompte, J.: Phase transition and complex systems, *Complexity* **1** (1996), 13–26.
8. Atlan, H.: *L'organisation biologique et la théorie de l'information*. Paris: Hermann, 1972.
9. Gatlin, L. L.: The information content of DNA, *J. of Theoret. Biol.* **10** (1966), 281–300.
10. Gatlin, L. L.: The information content of DNA II, *J. of Theoret. Biol.* **18** (1968), 181–194.
11. Skilling, J.: Theory of maximum entropy image reconstruction. In J.H. Justice (ed.), *Maximum entropy and Bayesian methods in applied statistics*. Cambridge University Press, UK, 1986: 156–178.

Address for correspondence: Guy Jumarie, Department of Mathematics, University of Quebec at Montreal, P.O. Box 8888, Downtown Street, Montreal, Qc., Canada H3C 3P8
 Phone: (514) 987-3000, ext. 7792; Fax: (514) 987-8935; e-mail: jumarie.guy@uqam.ca
 Phone (home): (514) 670-9981