Physics Letters A ••• (••••) •••–•••

Contents lists available at ScienceDirect

Physics Letters A

www.elsevier.com/locate/pla

Chaos forgets and remembers: Measuring information creation, destruction, and storage

Ryan G. James [a], Korana Burke [a], James P. Crutchfield [a,b,*]

[a] Complexity Sciences Center and Physics Department, University of California at Davis, One Shields Avenue, Davis, CA 95616, USA
[b] Santa Fe Institute, 1399 Hyde Park Road, Santa Fe, NM 87501, USA

ARTICLE INFO

Article history:
Received 21 September 2013
Received in revised form 1 April 2014
Accepted 15 May 2014
Available online xxxx
Communicated by C.R. Doering

Keywords:
Chaos
Entropy rate
Bound information
Shannon information measures
Information diagram
Discrete-time maps

ABSTRACT

The hallmark of deterministic chaos is that it creates information—the rate being given by the Kolmogorov–Sinai metric entropy. Since its introduction half a century ago, the metric entropy has been used as a unitary quantity to measure a system's intrinsic unpredictability. Here, we show that it naturally decomposes into two structurally meaningful components: A portion of the created information—the ephemeral information—is forgotten and a portion—the bound information—is remembered. The bound information is a new kind of intrinsic computation that differs fundamentally from information creation: it measures the rate of active information storage. We show that it can be directly and accurately calculated via symbolic dynamics, revealing a hitherto unknown richness in how dynamical systems compute.

The world is replete with systems that generate information—information that is then encoded in a variety of ways: Erratic ant behavior eventually leads to intricate, structured colony nests [1,2]; thermally fluctuating magnetic spins form complex domain structures [3]; music weaves theme, form, and melody with surprise and innovation [4]. We now appreciate that the underlying dynamics in such systems is frequently deterministic chaos [5,6]. In others, the underlying dynamics appears to be fundamentally stochastic [7]. For continuous-state systems, at least, one operational distinction between deterministic chaos and stochasticity is found in whether or not information generation diverges with measurement resolution [8]. This result calls back to Kolmogorov's original use [9] of Shannon's mathematical theory of communication [10] to measure a system's rate of information generation in terms of the metric entropy. Since that time, metric entropy has been understood as a unitary quantity. Whether deterministic or stochastic, it is a system's degree of unpredictability. Here, we show that this is far too simple a picture—one that obscures much.

To ground this claim, consider two systems. The first, a fair coin: Each flip is independent of the others, leading to a simple uncorrelated randomness. As a result, no statistical fluctuation is predictively informative. For the second system consider a stock traded via a financial market: While its price is unpredictable, the direction and magnitude of fluctuations can hint at its future behavior. (This, at least, is the guiding assumption of the now-global financial engineering industry.) We make this distinction rigorous here, dividing a system's information generation into a component that is relevant to temporal structure and a component divorced from it. We show that the structural component captures the system's internal information processing and, therefore, is of practical interest when harnessing the chaotic nature of physical systems to build novel machines and devices [11]. We first introduce the new measures, describe how to interpret and calculate them, and then apply them via a generating partition to analyze several dynamical systems—the Logistic, Tent, and Lozi maps—revealing a previously hidden form of active information storage.

We observe these systems via an optimal measuring instrument—called a generating partition—that encodes all of their behaviors in a *stationary process*: A distribution $\Pr(\ldots, X_{-2}, X_{-1}, X_0, X_1, X_2, \ldots)$ over a bi-infinite sequence of random variables with shift-invariant statistics. A contiguous block of observations $X_{t:t+\ell}$ begins at index t and extends for length ℓ. (The index is inclusive on the left and exclusive on the right.) If an index is infinite, we leave it blank. So, a process is compactly denoted $\Pr(X_:)$. Our analysis splits $X_:$ into three segments: the *present* X_0, a single observation; the *past* $X_{:0}$, everything prior; and *future* $X_{1:}$, everything that follows.

* Corresponding author.
E-mail addresses: rgjames@ucdavis.edu (R.G. James), kburke@ucdavis.edu (K. Burke), chaos@ucdavis.edu (J.P. Crutchfield).

http://dx.doi.org/10.1016/j.physleta.2014.05.014

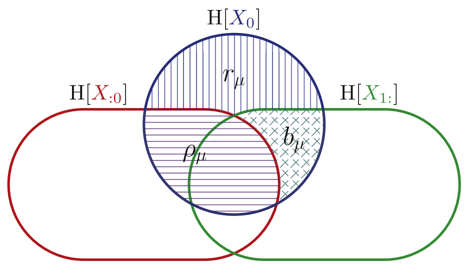

Fig. 1. A process's I-diagram showing how the past $X_{:0}$, present X_0, and future $X_{1:}$ partition each other into seven distinct information *atoms*. We focus only on the four regions contained in the *present information* H[X_0] (blue circle). That is, the present decomposes into three components: ρ_μ (horizontal lines), r_μ (vertical lines), and b_μ (diagonal crosshatching). The *redundant* information ρ_μ overlaps with the *past* H[$X_{:0}$]; the *ephemeral* information r_μ falls outside both the *past* and the *future* H[$X_{1:}$]. The *bound* information b_μ is that part of H[X_0] which is in the future yet not in the past. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

The information-theoretic relationships between these three random variable segments are graphically expressed in a Venn-like diagram, known as an I-diagram [12]; see Fig. 1. The rate h_μ of information generation is the amount of new information in an observation X_0 given all the prior observations $X_{:0}$:

$$h_\mu = \mathrm{H}[X_0 | X_{:0}], \tag{1}$$

where H[$Y|Z$] denotes the Shannon conditional entropy of random variable Y given variable Z. This quantity arises in various contexts and goes by many names: e.g., the Shannon entropy rate and the Kolmogorov–Sinai metric entropy, mentioned above [8]. The complement of the entropy rate is the *predicted information* ρ_μ:

$$\rho_\mu = \mathrm{I}[X_{:0} : X_0], \tag{2}$$

where I[$Y : Z$] denotes the mutual information between random variables Y and Z [12]. Hence, ρ_μ is the information in the present that can be predicted from prior observations. Together, we have a decomposition of the information contained in the present: $\mathrm{H}[X_0] = h_\mu + \rho_\mu$.

A simple application of the entropy chain rule [12] to Eq. (1) leads us to a different view:

$$\begin{aligned} h_\mu &= \mathrm{I}[X_0 : X_{1:} | X_{:0}] + \mathrm{H}[X_0 | X_{:0}, X_{1:}] \\ &= b_\mu + r_\mu. \end{aligned} \tag{3}$$

This introduces two new information measures:

$$b_\mu = \mathrm{I}[X_0 : X_{1:} | X_{:0}] \quad \text{and} \tag{4}$$

$$r_\mu = \mathrm{H}[X_0 | X_{:0}, X_{1:}]. \tag{5}$$

That is, created information (h_μ) decomposes into two parts: information (b_μ) shared by the present and the future but not in the past and information (r_μ) in the present but in neither the past nor the future.

The r_μ component was first studied by Verdú and Weissman [13] as the *erasure entropy* (their H^-) to measure information loss in erasure channels. To emphasize that it is information existing only in a single moment—created and then immediately forgotten—we refer to r_μ as the *ephemeral information*. The second component b_μ we call the *bound information* since it is information created in the present that the system stores and that goes on to affect the future.[1] It was first studied as a measure of "interestingness" in computational musicology by Abdallah and Plumbley [14]. For a more complete analysis of this decomposition, as well as computation methods and related measures, see Ref. [15].

Isolating the information H[X_0] contained in the present and identifying its components provides the partitioning illustrated in Fig. 1. This is a particularly intuitive way of thinking about the information contained in an observation. While some behavior (ρ_μ) can be predicted, the rest ($h_\mu = b_\mu + r_\mu$) cannot. Of that which cannot be predicted, some (b_μ) plays a role in the future behavior and some (r_μ) does not. As such, this is a natural decomposition of a time series; one that results in a semantic dissection of the entropy rate.

By way of an example, consider a few simple processes and how their present information decomposes into these three components. A periodic process of alternating 0s and 1s (…01010101…) has H[X_0] = 1 bit since 0s and 1s occur equally often. Given a prior observation, one can accurately predict exactly which symbol will occur next and so $\mathrm{H}[X_0] = \rho_\mu = 1$ bit, while $r_\mu = b_\mu = 0$ bits. On the other extreme is a fair coin flip. Again, each outcome is equally likely and so H[X_0] = 1 bit. However, each flip is independent of all others and so $\mathrm{H}[X_0] = r_\mu = 1$ bit, while $\rho_\mu = b_\mu = 0$ bits.

Between these two extrema lie interesting processes: those with *stochastic structure*. Processes expressing a fixed template, like the periodic process above, contain a finite amount of information. Those with stochastic structure, however, constantly generate information and store it in the form of patterns. Being neither purely predictable nor independently random, these patterns are captured by b_μ. The more intricate the organization, the larger b_μ. More to the point, generating these patterns requires intrinsic computation in a system—information creation, storage, and transformation [16]. We propose b_μ as a simple method of discovering this type of physical computation: Where there are intricate patterns, there is sophisticated processing.

How useful is the proposed decomposition and its measures? To answer this we analyze several discrete-time chaotic dynamical systems—the Logistic and Tent maps of the interval and the Lozi map of the plane—uncovering a number of novel properties embedded in these familiar and oft-studied systems. As an independent calibration for the measures, we employ Pesin's theorem [17]: h_μ is the sum of the positive Lyapunov characteristic exponents (LCEs). The maps here have at most one positive LCE λ, so $h_\mu = \max\{0, \lambda\}$. The symbols $s_0, s_1, s_2, \ldots, s_N$ for each process we analyze come from a generating partition. We produce a long sample of $N \approx 10^{10}$ symbols, extracting subsequence statistics via a sliding window.[2] Each window consists of a past, present, and future symbol sequence and we estimate r_μ and b_μ using truncated forms of Eqs. (4) and (5).

Consider first the Logistic map, perhaps one of the most studied chaotic systems:

$$x_{n+1} = a x_n (1 - x_n), \tag{6}$$

where $a \in [0, 4]$ is the control parameter and the initial condition is $x_0 \in [0, 1]$. Its generating partition is defined by:

$$s_n = \begin{cases} 0 & \text{if } x_n < \frac{1}{2}, \\ 1 & \text{if } x_n \geq \frac{1}{2}. \end{cases} \tag{7}$$

[1] Our terminology avoids the misleading use of the phrase "predictive information" for b_μ. The latter is not the amount of information needed to predict the future. Rather, it is part of the *predictable* information—that portion of the future which *can* be predicted.

[2] Window width is adaptively chosen in inverse proportion to the LCE. When the latter is low we use a longer window than when the system is fully chaotic. The minimum window width of $L = 31$ and adaptive widths were chosen so that numerical estimates varied by less than 0.01% when the width is incremented.