Informational and Causal Architecture of Continuous-time Renewal and Hidden Semi-Markov Processes

Sarah E. Marzen
James P. Crutchfield

SFI WORKING PAPER: 2016-11-024

Santa Fe Institute Working Paper 16-10-XXX
arxiv.org:1610.01099 [cond-mat.stat-mech]

Informational and Causal Architecture of Continuous-time Renewal and Hidden Semi-Markov Processes

Sarah Marzen[1, 2, *] and James P. Crutchfield[3, †]

[1] *Physics of Living Systems Group, Department of Physics, Massachusetts Institute of Technology, Cambridge, MA 02139*
[2] *Department of Physics, University of California at Berkeley, Berkeley, CA 94720*
[3] *Complexity Sciences Center and Department of Physics, University of California at Davis, One Shields Avenue, Davis, CA 95616*

(Dated: November 4, 2016)

We introduce the minimal maximally predictive models (ϵ-machines) of processes generated by certain hidden semi-Markov models. Their causal states are either hybrid discrete-continuous or continuous random variables and causal-state transitions are described by partial differential equations. Closed-form expressions are given for statistical complexities, excess entropies, and differential information anatomy rates. We present a complete analysis of the ϵ-machines of continuous-time renewal processes and, then, extend this to processes generated by unifilar hidden semi-Markov models and semi-Markov models. Our information-theoretic analysis leads to new expressions for the entropy rate and the rates of related information measures for these very general continuous-time process classes.

PACS numbers: 02.50.-r 89.70.+c 05.45.Tp 02.50.Ey 02.50.Ga
Keywords: renewal process, entropy rate, excess entropy, statistical complexity, information anatomy, hidden semi-Markov models

I. INTRODUCTION

We are interested in answering two very basic questions about continuous-time, discrete-symbol stochastic processes:

- What are their minimal maximally predictive models—their ϵ-machines?
- What are information-theoretic characterizations of their randomness, predictability, and complexity?

For shorthand, we refer to the former as *causal architecture* and the latter as *informational architecture*. Minimal maximally predictive models of discrete-time, discrete-state, discrete-output processes are relatively well understood; e.g., see Refs. [1–3]. Some progress has been made on understanding minimal maximally predictive models of discrete-time, continuous-output processes; e.g., see Refs. [4–6]. Relatively less is understood about minimal maximally predictive models of continuous-time, discrete-output processes, beyond those with exponentially decaying state-dwell times [6]. The following is a first attempt at a remedy that complements the spectral methods developed in Ref. [6], as we address the less tractable case of uncountably infinite causal states.

We start by analyzing continuous-time renewal processes, as addressing the challenges there carries over to other continuous-time processes. (Elsewhere, we outline the wide interest and applicability of renewal processes in physics and the quantitative sciences generally [7–9].) The difficulties are both technical and conceptual. First, the causal states are now continuous or hybrid discrete-continuous random variables, unless the renewal process is Poisson. Second, transitions between causal states are now described by partial differential equations. Finally, and perhaps most challenging, most informational architecture quantities must be redefined. With these challenges addressed, we turn our attention to a very general class of continuous-time, discrete-alphabet processes—stateful renewal processes generated by unifilar hidden semi-Markov models. We identify their ϵ-machines and find new expressions for entropy rate and other informational architecture quantities, extending results in Ref. [10].

Our main thesis is rather simple: minimal maximally predictive models of continuous-time, discrete-symbol processes require a wholly new ϵ-machine calculus. To develop it, Sec. II describes the required new notation and definitions that enable extending the ϵ-machine framework which is otherwise well understood for discrete-time processes [1, 11]. Sections III-V determine the causal and informational architecture of continuous-time renewal processes. Section VI characterizes the ϵ-machines and calculates the entropy rate and excess entropy of unifilar hidden semi-Markov models. We conclude by describing potential applications to Bayesian ϵ-machine in-

* semarzen@mit.edu
† chaos@ucdavis.edu

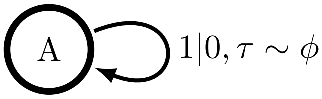

FIG. 1. Generative model of a continuous-time renewal process. The length $\mathcal{T}_i = \tau$ of periods of silence (corresponding to output symbol 0) are drawn independently, identically distributed (IID) from probability density $\phi(t)$.

ference algorithms using new enumerations of ϵ-machine topologies and to information measure estimation using the formulae of Sec. VI.

II. BACKGROUND AND NOTATION

A continuous-time, discrete-symbol time series $\ldots, (x_{-1}, \tau_{-1}), (x_0, \tau_0), (x_1, \tau_1), \ldots$ is described by a list of symbols x in a finite alphabet $\mathcal{A}$ and dwell times $\tau \in \Re_{\geq 0}$ for those symbols. In this representation, we demand that $x_i \neq x_{i+1}$ to enforce a unique presentation of the time series.

Sections III-V focus on point processes for which $|\mathcal{A}| = 1$. And so, in this case, we label the time series only with dwell times: $\ldots, \tau_{-1}, \tau_0, \tau_1, \ldots$. We view the time series $\overleftrightarrow{\tau}$ as a realization of random variables $\overleftrightarrow{\mathcal{T}} = \ldots, \mathcal{T}_{-1}, \mathcal{T}_0, \mathcal{T}_1, \ldots$. When the observed time series is strictly stationary and the process ergodic, in principle, we can calculate the probability distribution $\Pr(\overleftrightarrow{\mathcal{T}})$ from a single realization $\overleftrightarrow{\tau}$.

Demarcating the present splits $\mathcal{T}_0$ into two parts: the time $\mathcal{T}_{0^+}$ since first emitting the previous symbol and the time $\mathcal{T}_{0^-}$ to next symbol. Thus, we define $\mathcal{T}_{-\infty:0^+} = \ldots, \mathcal{T}_{-1}, \mathcal{T}_{0^+}$ as the *past* and $\mathcal{T}_{0^-:\infty} = \mathcal{T}_{0^-}, \mathcal{T}_1, \ldots$ as the *future*. (To reduce notation, we drop the ∞ indices.) The *present* $\mathcal{T}_{0^+:0^-}$ itself extends over an infinitesimally small length of time.

Continuous-time renewal processes have a relatively simple generative model. *Interevent intervals* $\mathcal{T}_i$ are drawn from a probability density function $\phi(t)$. The *survival function* $\Phi(t) = \int_t^\infty \phi(t')dt'$ is the probability that an interevent interval is greater than or equal to t and, in a nod to neuroscience, we define the *mean firing rate* μ as:

$$\mu^{-1} = \int_0^\infty t\phi(t)dt \ .$$

The minimal generative model for a continuous-time renewal process is therefore a single causal-state machine with a continuous-value observable $\mathcal{T}$; as shown in Fig. 1.

A. Causal architecture

A process' *forward-time causal states* are defined, as usual, by the *predictive* equivalence relation [1], written here for the case of point processes:

$$\tau_{:0^+} \sim_{\epsilon^+} \tau_{:0^+}{}' \Leftrightarrow \Pr(\mathcal{T}_{0^-:}|\mathcal{T}_{:0^+} = \tau_{:0^+}) = \Pr(\mathcal{T}_{0^-:}|\mathcal{T}_{:0^+} = \tau_{:0^+}{}') \ .$$

It is straightforward to write the predictive equivalence relation for continuous-time, discrete-alphabet point processes using the notation. This partitions the set of allowed pasts. Each equivalence class of pasts is a forward-time causal state $\sigma^+ = \epsilon^+(\tau_{:0^+})$, in which $\epsilon^+(\cdot)$ is the function that maps a past to its causal state. The set of forward-time causal states $\mathcal{S}^+ = \{\sigma^+\}$ inherits a probability distribution $\Pr(\mathcal{S}^+)$ from the probability distribution over pasts $\Pr(\mathcal{T}_{:0^+})$. *Forward-time prescient statistics* are any refinement of the forward-time causal-state partition. By construction, they are a sufficient statistic for prediction, but not necessarily *minimal* sufficient statistics [1].

Reverse-time causal states are essentially forward-time causal states of the time-reversed process. In short, reverse-time causal states $\mathcal{S}^- = \{\sigma^-\}$ are the classes defined by the retrodictive equivalence relation, written here for the case of point processes:

$$\tau_{0^-:} \sim_{\epsilon^-} \tau'_{0^-:} \Leftrightarrow \Pr(\mathcal{T}_{:0^+}|\mathcal{T}_{0^-:} = \tau_{0^-:}) = \Pr(\mathcal{T}_{:0^+}|\mathcal{T}_{0^-:} = \tau'_{0^-:}) \ .$$

It is, again, straightforward to write the predictive equivalence relation for continuous-time, discrete-alphabet point processes using the notation given above. And, similarly, reverse-time causal states $\mathcal{S}^- = \epsilon^-(\mathcal{T}_{0^-:})$ inherit a probability measure $\Pr(\mathcal{S}^-)$ from the probability distribution $\Pr(\mathcal{T}_{0^-:})$ over futures. Reverse-time prescient statistics are any refinement of the reverse-time causal-state partition. They are sufficient statistics for retrodiction, but not necessarily minimal.

The main import of these definitions derives from the *causal shielding* relations:

$$\Pr(\mathcal{T}_{0^-:}, \mathcal{T}_{:0^+}|\mathcal{S}^+) = \Pr(\mathcal{T}_{0^-:}|\mathcal{S}^+) \Pr(\mathcal{T}_{:0^+}|\mathcal{S}^+) \quad (1)$$

$$\Pr(\mathcal{T}_{0^-:}, \mathcal{T}_{:0^+}|\mathcal{S}^-) = \Pr(\mathcal{T}_{0^-:}|\mathcal{S}^-) \Pr(\mathcal{T}_{:0^+}|\mathcal{S}^-) \ . \quad (2)$$

The consequence of these is illustrated in Fig. 2. That is, arbitrary functions of the past and future do not shield the two aggregate past and future random variables from one another. So, these causal shielding relations are special to prescient statistics, causal states, and their defining functions $\epsilon^+(\cdot)$ and $\epsilon^-(\cdot)$. Forward and reverse-time generative models do not, in general, have state spaces that satisfy Eqs. (1) and (2).

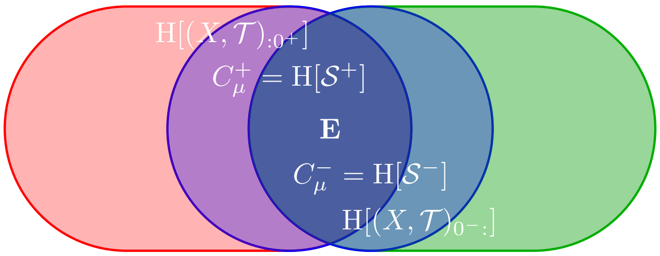

FIG. 2. Predictability, compressibility, and causal irreversibility in renewal and semi-Markov processes graphically illustrated using a Venn-like *information diagram* over the random variables for the past $(X,\mathcal{T})_{:0^+}$ (left oval, red), the future $(X,\mathcal{T})_{0^-:}$ (right oval, green), the forward-time causal states $\mathcal{S}^+$ (left circle, purple), and the reverse-time causal states $\mathcal{S}^-$ (right circle, blue). (Cf. Ref. [12].) The forward-time and reverse-time statistical complexities are the entropies of $\mathcal{S}^+$ and $\mathcal{S}^-$, i.e., the memories required to losslessly predict or retrodict, respectively. The excess entropy $\mathbf{E} = \mathrm{I}[(X,\mathcal{T})_{:0^+};(X,\mathcal{T})_{0^-:}]$ is a measure of process predictability (central pointed ellipse, dark blue) and Theorem 1 of Ref. [12, 13] shows that $\mathbf{E} = \mathrm{I}[\mathcal{S}^+;\mathcal{S}^-]$ by applying the causal shielding relations in Eqs. (1) and (2).

The *forward-time ϵ-machine* is that with state space $\mathcal{S}^+$ and transition dynamic between forward-time causal states. The *reverse-time ϵ-machine* is that with state space $\mathcal{S}^-$ and transition dynamic between reverse-time causal states. Defining these transition dynamics for continuous-time processes requires a surprising amount of care, as discussed in Secs. III-V.

B. Informational architecture

We are broadly interested in information-theoretic characterizations of a process' predictability, compressibility, and randomness. A list of current quantities of interest, though by no means exhaustive, is given in Figs. 2 and 6. Curiously, many lose meaning when naively applied to continuous-time processes; e.g., see Refs. [5, 9, 14]. This section, as a necessity, will redefine many of these in relatively simple, but new ways to avoid trivial divergences and zeros.

The *forward-time statistical complexity* $C_\mu^+ = \mathrm{H}[\mathcal{S}^+]$ is the cost of coding the forward-time causal states and the *reverse-time statistical complexity* $C_\mu^- = \mathrm{H}[\mathcal{S}^-]$ is the cost of coding reverse-time causal states. When $\mathcal{S}^+$ or $\mathcal{S}^-$ are mixed or continuous random variables, one employs differential entropies for $\mathrm{H}[\cdot]$. The result, though, is that the statistical complexities are potentially negative or infinite or both [15, Ch. 8.3], perhaps undesirable characteristics for a definition of process complexity. This definition, however, allows for consistency with complexity definitions for discretized continuous-time processes. See Ref. [16] for possible alternatives for $\mathrm{H}[\cdot]$.

Together, a process' *causal irreversibility* [12, 13] is defined as the difference between the forward and reverse-time statistical complexities:

$$\Xi = C_\mu^+ - C_\mu^- .$$

If the forward- and reverse-time process ϵ-machines are isomorphic—i.e., if the process is temporally reversible—then $\Xi = 0$.

III. CONTINUOUS-TIME CAUSAL STATES

Renewal processes are temporally symmetric: $\Xi = 0$ [7]. As such, we will refer to forward-time causal states and the forward-time ϵ-machine as simply causal states or the ϵ-machine, with the understanding that reverse-time causal states and reverse-time ϵ-machines will take the exact same form with slight labeling differences.

We start by describing prescient statistics for continuous-time processes. The Lemma which does this exactly parallels that of Lemma 1 of Ref. [7]. The only difference is that the prescient statistic is the *time* since last event, rather than the number of 0s (count) since last event.

Lemma 1. *The time $\mathcal{T}_{0^+}$ since last event is a prescient statistic of renewal processes.*

Proof. *From Bayes Rule:*

$$\Pr(\mathcal{T}_{0^-:}|\mathcal{T}_{:0^+}) = \Pr(\mathcal{T}_{0^-}|\mathcal{T}_{0^+:})\Pr(\mathcal{T}_{1:}|\mathcal{T}_{:1}) .$$

Interevent intervals $\mathcal{T}_i$ are independent of one another, so $\Pr(\mathcal{T}_{1:}|\mathcal{T}_{:1}) = \Pr(\mathcal{T}_{1:})$. The random variables $\mathcal{T}_{0^+}$ and $\mathcal{T}_{0^-}$ are functions of $\mathcal{T}_0$ and the location of the present. Both $\mathcal{T}_{0^+}$ and $\mathcal{T}_{0^-}$ are independent of other interevent intervals. And so, $\Pr(\mathcal{T}_{0^-}|\mathcal{T}_{0^+:}) = \Pr(\mathcal{T}_{0^-}|\mathcal{T}_{0^+})$. This implies:

$$\Pr(\mathcal{T}_{0^-:}|\mathcal{T}_{:0^+}) = \Pr(\mathcal{T}_{1:})\Pr(\mathcal{T}_{0^-}|\mathcal{T}_{0^+}) . \quad (3)$$

The predictive equivalence relation groups two pasts $\tau_{:0^+}$ and $\tau'_{:0^+}$ together when $\Pr(\mathcal{T}_{0^-:}|\mathcal{T}_{:0^+} = \tau_{:0^+}) = \Pr(\mathcal{T}_{0^-:}|\mathcal{T}_{:0^+} = \tau'_{:0^+})$. We see that $\tau_{0^+} = \tau'_{0^+}$ is a sufficient condition for this from Eq. (3). The Lemma follows.

Some renewal processes are quite predictable, while others are purely random. A Poisson process is the latter: Interevent intervals are drawn independently from an exponential distribution and so knowing the time since last event provides no predictive benefit. A fractal renewal process can be the former. There, the interevent interval is so structured that the resultant process can have power-law correlations [17]. Then, knowing the time since last event can provide quite a bit of predictive power [8].

Intermediate between these two extremes is a broad class of renewal processes whose interevent intervals are structured up to a point and then fall off exponentially only after some time T^*. These intermediate cases can be classified as either of the following types of renewal process, in analogy with Ref. [7]'s classification. Note that an eventually Δ-Poisson process, but not an eventually Poisson process, will generally have a discontinuous $\phi(t)$.

Definition 1. *An* eventually Poisson process *has:*

$$\phi(t) = \phi(T)e^{-\lambda(t-T)} ,$$

for some $\lambda > 0$ and $T > 0$ almost everywhere. We associate the eventually Poisson process with the minimal such T.

Definition 2. *An* eventually Δ-Poisson process *with $\Delta^* > 0$ has an interevent interval distribution satisfying:*

$$\phi(T^* + s) = \phi(T^* + (s - T^*) \bmod \Delta^*)e^{-\lambda \lfloor s/\Delta^* \rfloor}$$

for the smallest possible T^* for which Δ^* exists.

A familiar example of an eventually Poisson process is found in the spike trains generated by Poisson neurons with refractory periods [9]. There, the neuron is effectively prevented from firing two spikes within a time T of each other—the period during with its ion channels re-energize the membrane voltage to their nonequilibrium steady state. After that, the time to next spike is drawn from an exponential distribution. To exactly predict the spike train's future, we must know the time since last spike, as long as it is less than T. We gain a great deal of predictive power from that piece of information. However, we do not care much about the time since last spike exactly if it is greater than T, since at that point the neuron acts as a memoryless Poisson neuron. These intuitions are captured by the following classification theorem.

Theorem 1. *A renewal process has three different types of causal state:*

1. *When the renewal process is not eventually Δ-Poisson, the causal states are the time since last event;*
2. *When the renewal process is eventually Poisson, the causal states are the time since last event up until time T^*; or*
3. *When the renewal process is eventually Δ-Poisson, the causal states are the time since last event up until time T^* and are the times since T^* mod Δ thereafter.*

Proof. *Lemma 1 implies that two pasts are causally equivalent if they have the same time since last event, if $\tau_{0^+} = \tau'_{0^+}$. From Lemma 1's proof, we further see that two times since last event are causally equivalent when $\Pr(\mathcal{T}_{0^-}|\mathcal{T}_{0^+} = \tau_{0^+}) = \Pr(\mathcal{T}_{0^-}|\mathcal{T}_{0^+} = \tau'_{0^+})$. In terms of $\phi(t)$, we find that:*

$$\Pr(\mathcal{T}_{0^-} = \tau_{0^-}|\mathcal{T}_{0^+} = \tau_{0^+}) = \frac{\phi(\tau_{0^-} + \tau_{0^+})}{\Phi(\tau_{0^+})} ,$$

using manipulations very similar to those in the proof of Thm. 1 of Ref. [7]. So, to find causal states, we look for $\tau_{0^+} \neq \tau'_{0^+}$ such that:

$$\frac{\phi(\tau_{0^-} + \tau_{0^+})}{\Phi(\tau_{0^+})} = \frac{\phi(\tau_{0^-} + \tau'_{0^+})}{\Phi(\tau'_{0^+})} .$$

for all $\tau_{0^-} \geq 0$.

To unravel the consequences of this, we suppose that $\tau_{0^+} < \tau'_{0^+}$ without loss of generality. Define $\Delta = \tau'_{0^+} - \tau_{0^+}$ and $T = \tau_{0^+}$, for convenience. The predictive equivalence relation can then be rewritten as:

$$\phi(T + \Delta + \tau_{0^-}) = \lambda\phi(T + \tau_{0^-}) ,$$

for any $\tau_{0^-} \geq 0$, where $\lambda = \Phi(T + \Delta)/\Phi(T)$. Iterating this relationship, we find that:

$$\phi(T + \tau_{0^-}) = \lambda^{\lfloor \tau_{0^-}/\Delta \rfloor}\phi\left(T + (\tau_{0^-} \bmod \Delta)\right) .$$

This immediately implies the theorem's first case. If a renewal process is not *eventually Δ-Poisson, then $\phi(\tau_{0^-} + \tau_{0^+})/\Phi(\tau_{0^+}) = \phi(\tau_{0^-} + \tau'_{0^+})/\Phi(\tau'_{0^+})$ for all $\tau_{0^-} \geq 0$ implies $\tau_{0^+} = \tau'_{0^+}$, so that the prescient statistics of Lemma 1 are also minimal.*

To understand the theorem's last two cases, we consider more carefully the set of all pairs (T, Δ) for which $\phi(\tau_{0^-} + T)/\Phi(T) = \phi(\tau_{0^-} + T + \Delta)/\Phi(T + \Delta)$ for all $\tau_{0^-} \geq 0$ holds. Define the set:

$$\mathcal{S}_{T,\Delta} := \Big\{(T, \Delta) : \frac{\phi(\tau_{0^-} + T)}{\Phi(T)} = \frac{\phi(\tau_{0^-} + T + \Delta)}{\Phi(T + \Delta)}, \text{ for all } \tau_{0^-} \geq 0\Big\}$$

and define the parameters T^* and Δ^* by:

$$T^* := \inf\{T : \text{ there exists } \Delta \text{ such that } (T, \Delta) \in \mathcal{S}_{T,\Delta}\}$$

and:

$$\Delta^* := \inf\{\Delta : (T^*, \Delta) \in \mathcal{S}_{T,\Delta}\} .$$

Note that T^* and Δ^* defined in this way are unique and exist, as we assumed that $\mathcal{S}_{T,\Delta}$ is nonempty. When $\Delta^* > 0$, then the process is eventually Δ-Poisson. If $\Delta^* = 0$, then the process must be an eventually Poisson process with parameter T^*. To see this, we return to the equation:

$$\phi(T^* + \Delta + \tau_{0-}) = \frac{\Phi(T^* + \Delta)}{\Phi(T^*)} \phi(T^* + \tau_{0-}) \ ,$$

and rearrange terms to find:

$$\frac{\phi(T^* + \Delta + \tau_{0-}) - \phi(T^* + \tau_{0-})}{\phi(T^* + \tau_{0-})} = \frac{\Phi(T^* + \Delta) - \Phi(T^*)}{\Phi(T^*)} .$$

As $\Delta^* = 0$, we can take the limit that $\Delta \to 0$ and we find that:

$$\frac{d \log \phi(t)}{dt}\Big|_{t=T^*+\tau_{0-}} = \frac{d \log \Phi(t)}{dt}\Big|_{t=T^*} \ .$$

The righthand side is a parameter independent of τ_{0-}. So, this is a standard ordinary differential equation for $\phi(t)$. It is solved by $\phi(t) = \phi(T^*) e^{-\lambda(t-T^*)}$ for $\lambda := -d \log \Phi(t)/dt\big|_{t=T^*}$.

Theorem 1 implies that there is a qualitative change in $\mathcal{S}^+$ depending on whether or not the renewal process is Poisson, eventually Poisson, eventually Δ-Poisson, or not eventually Poisson. In the first case, $\mathcal{S}^+$ is a discrete random variable; in the second case, $\mathcal{S}^+$ is a mixed discrete-continuous random variable; and in the third and fourth cases, $\mathcal{S}^+$ is a continuous random variable.

IV. WAVE PROPAGATION ON CONTINUOUS-TIME ϵ-MACHINES

Identifying causal states in continuous-time follows an almost entirely similar path to that used for discrete-time renewal processes in Ref. [7]. The seemingly slight differences between the causal states of eventually Poisson, eventually Δ-Poisson, and not eventually Δ-Poisson renewal processes, however, have surprisingly important consequences for continuous-time ϵ-machines.

As described by Thm. 1, there are often an uncountable infinity of continuous-time causal states. As one might anticipate from Refs. [7, 9], however, there is an ordering to this infinity of causal states that makes calculations tractable. There is one major difference between discrete-time ϵ-machines and continuous-time ϵ-machines: transition dynamics often amount to specifying the evolution of a probability density function over causal-state space.

As such, a continuous-time ϵ-machine constitutes an unusual presentation of a hidden Markov model: they appear as a system of conveyor belts or, under special conditions, like conveyor belts with a trash bin or a second mini-conveyor belt. Beyond the picaresque metaphor, in fact they operate like conveyor belts in that they transport the time since the last event, resetting it here and there in a stateful way.

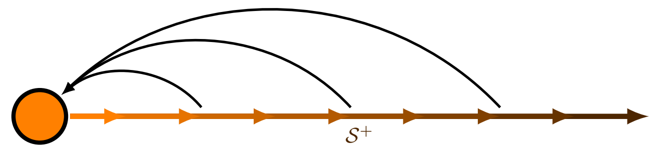

FIG. 3. ϵ-Machine for the generic not eventually Poisson renewal process: Continuous-time causal states $\mathcal{S}^+$, tracking the time since last event and depicted as the semi-infinite horizontal line, are isomorphic with the positive real line. If no event is seen, probability flows towards increasing time since last event, as described in Eq. (6). Otherwise, arrows denote allowed transitions back to the reset state or "0 node" (solid black circle at left), denoting that an event occurred.

Unsurprisingly, the exception to this general rule is given by the Poisson process itself. The ϵ-machine of a Poisson process is exactly the minimal generative model shown in Fig. 1. At each iteration, an interevent interval is drawn from a probability density function $\phi(t) = \lambda e^{-\lambda t}$, with $\lambda > 0$. Knowing the time since last event does not aid in predicting the time to next event, above and beyond knowing λ. And so, the Poisson ϵ-machine has only a single state.

In the general setting, though, the ϵ-machine dynamic describes the evolution of the probability density function over its causal states. How to represent this? We might search for labeled transition operators $\mathcal{O}^{(x)}$ such that $\partial \rho(\sigma, t)/\partial t = \mathcal{O}^{(x)} \rho(\sigma, t)$, giving partial differential equations that govern the labeled-transition dynamics.

A. Not Eventually Poisson

The ϵ-machine of a renewal process that is not eventually Poisson takes the state-transition form shown in Fig. 3. Let $\rho(\sigma, t)$ be the probability density function over the causal states σ at time t. Our approach to deriving labeled transition dynamics parallels well-known approaches to determining Fokker-Planck equations using a Kramers-Moyal expansion [18]. Here, this means that any probability at causal state σ at time $t + \Delta t$ could only have come from causal state $\sigma - \Delta t$ at time t, if $\sigma \geq \Delta t$. This implies:

$$\begin{aligned} \rho(\sigma, & t + \Delta t) \\ & = \Pr(\mathcal{S}_{t+\Delta t} = \sigma | \mathcal{S}_t = \sigma - \Delta t) \rho(\sigma - \Delta t, t) \ . \end{aligned} \quad (4)$$

However, $\Pr(\mathcal{S}_{t+\Delta t} = \sigma | \mathcal{S}_t = \sigma - \Delta t)$ is simply the probability that the interevent interval is greater than σ, given that the interevent interval is at least $\sigma - \Delta t$, or:

$$\Pr(\mathcal{S}_{t+\Delta t} = \sigma | \mathcal{S}_t = \sigma - \Delta t) = \frac{\Phi(\sigma)}{\Phi(\sigma - \Delta t)} \ . \tag{5}$$

Together, Eqs. (4) and (5) imply that:

$$\rho(\sigma, t + \Delta t) = \frac{\Phi(\sigma)}{\Phi(\sigma - \Delta t)} \rho(\sigma - \Delta t, t) \ .$$

From this, we obtain:

$$\begin{aligned}\frac{\partial \rho(\sigma,t)}{\partial t} &= \lim_{\Delta t \to 0} \frac{\rho(\sigma, t+\Delta t) - \rho(\sigma,t)}{\Delta t} \\ &= \lim_{\Delta t \to 0} \frac{\frac{\Phi(\sigma)}{\Phi(\sigma-\Delta t)} \rho(\sigma - \Delta t, t) - \rho(\sigma, t)}{\Delta t} \\ &= \lim_{\Delta t \to 0} \frac{(\frac{\Phi(\sigma)}{\Phi(\sigma-\Delta t)} - 1)\rho(\sigma - \Delta t, t)}{\Delta t} \\ &\quad + \lim_{\Delta t \to 0} \frac{\rho(\sigma - \Delta t, t) - \rho(\sigma, t)}{\Delta t} \\ &= \frac{\partial \log \Phi(\sigma)}{\partial \sigma} \rho(\sigma, t) - \frac{\partial \rho(\sigma, t)}{\partial \sigma} \ . \end{aligned} \tag{6}$$

Hence, the labeled transition operator $\mathcal{O}^{(0)}$ given no event takes the form:

$$\mathcal{O}^{(0)} = \frac{\partial \log \Phi(\sigma)}{\partial \sigma} - \frac{\partial}{\partial \sigma} \ .$$

The probability density function $\rho(\sigma, t)$ changes discontinuously after an event occurs, though. All probability mass shifts from $\sigma > 0$ resetting back to $\sigma = 0$:

$$\mathcal{O}^{(1)} \rho(\sigma, t) = -\frac{\phi(\sigma)}{\Phi(\sigma)} \rho(\sigma, t) + \delta(\sigma) \int_0^\infty \frac{\phi(\sigma')}{\Phi(\sigma')} \rho(\sigma', t) d\sigma' \ .$$

In other words, an event "collapses the wavefunction".

The stationary distribution $\rho(\sigma)$ over causal states is given by setting $\partial_t \rho(\sigma, t)$ to 0 and solving. (At the risk of notational confusion, we adopt the convention that $\rho(\sigma)$ denotes the stationary distribution and that $\rho(\sigma, t)$ does not.) Straightforward algebra shows that:

$$\rho(\sigma) = \mu \Phi(\sigma) \ .$$

From this, the continuous-time statistical complexity directly follows:

$$C_\mu = \int_0^\infty \mu \Phi(\sigma) \log \frac{1}{\mu \Phi(\sigma)} d\sigma \ .$$

This was the nondivergent component of the infinitesimal time-discretized renewal process' statistical complexity found in Ref. [9].

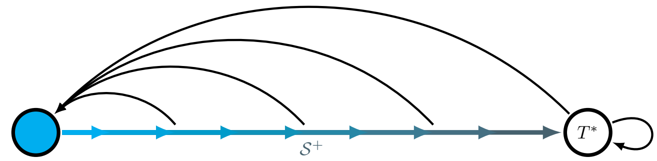

FIG. 4. ϵ-Machine for an eventually Poisson renewal process: Continuous-time causal states $\mathcal{S}^+$ are isomorphic with the real line only to $[0, T^*]$, as they again denote time since last event. A leaky absorbing node at T^* (solid white circle at right) corresponds to any time since last event after T^*. If no event is seen, probability flows towards increasing time since last event or the leaky absorbing node, as described in Eqs. (6) and (7). When an event occurs the process transitions (curved arrows) back to the reset state—node 0 (solid black circle at left).

B. Eventually Poisson

As Thm. 1 anticipates, there is a qualitatively different topology to the ϵ-machine of an eventually Poisson renewal process, largely due to the continuous-time causal states being mixed discrete-continuous random variables. For $\sigma < T^*$, there is "wave" propagation completely analogous to that described in Eq. (6) of Sec. IV A. However, there is a new kind of continuous-time causal state at $\sigma = T^*$, which does not have a one-to-one correspondence to the dwell time. Instead, it denotes that the dwell time is *at least* some value; viz., T^*. New notation follows accordingly: $\rho(\sigma, t)$, defined for $\sigma < T^*$, denotes a probability density function for $\sigma < T^*$ and $\pi(T^*, t)$ denotes the probability of existing in causal state $\sigma = T^*$. Normalization, then, requires that:

$$\int_0^{T^*} \rho(\sigma, t) d\sigma + \pi(T^*, t) = 1 \ .$$

The transition dynamics for $\pi(T^*, t)$ are obtained similarly to that for $\rho(\sigma, t)$, in that we consider all ways in which probability flows to $\pi(T^*, t + \Delta t)$ in a short time window Δt. Probability can flow from any causal state with $T^* - \Delta t \leq \sigma < T^*$ or from $\sigma = T^*$ itself. That is, if no event is observed, we have:

$$\begin{aligned}\pi(T^*, t + \Delta t) &= e^{-\lambda \Delta t} \pi(T^*, t) \\ &\quad + \int_{0^+}^{\Delta t} \rho(T^* - t', t) \frac{\Phi(T^*) e^{-\lambda(\Delta t - t')}}{\Phi(T^* - t')} dt'.\end{aligned}$$

The term $e^{-\lambda \Delta t} \pi(T^*, t)$ corresponds to probability flow from $\sigma = T^*$ and the integrand corresponds to probability influx from states $\sigma = T^* - t'$ with $0 < t' \leq \Delta t$. Assuming differentiability of $\pi(T^*, t)$ with respect to t,

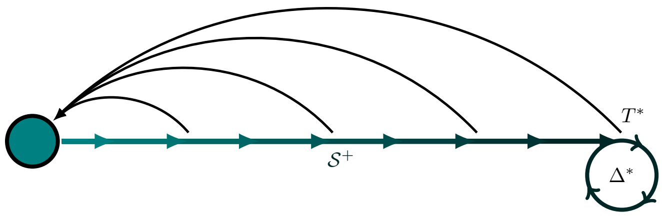

FIG. 5. ϵ-Machine for an eventually Δ-Poisson renewal process: Graphical elements as in the previous figure. The circular causal-state space at T^* (circle on right) has total duration Δ^*, corresponding to any time since last event after T^* mod Δ^*. If no event is seen, probability flows as indicated around the circle, as described in Eq. (6).

we find that:

$$\frac{\partial}{\partial t}\pi(T^*,t) = -\lambda\pi(T^*,t) + \rho(T^*,t) \ , \tag{7}$$

where $\rho(T^*,t)$ is shorthand for $\lim_{\sigma\to T^*}\rho(\sigma,t)$. This implies that the labeled transition operator $\mathcal{O}^{(0)}$ takes a piecewise form which acts as in Eq. (6) for $\sigma < T^*$ and as in Eq. (7) for $\sigma = T^*$. As earlier, observing an event causes the "wavefunction collapse" to a delta distribution at $\sigma = 0$.

The causal-state stationary distribution is determined again by setting $\partial_t\rho(\sigma,t)$ and $\partial_t\pi(\sigma,t)$ to 0. Equivalently, one can use the prescription suggested by Thm. 1 to calculate $\pi(T^*)$ via integration of the stationary distribution over the prescient machine given in Sec. IV A:

$$\begin{aligned}\pi(T^*) &= \int_{T^*}^{\infty}\rho(\sigma)d\sigma \\ &= \mu\int_{T^*}^{\infty}\Phi(\sigma)d\sigma \ .\end{aligned}$$

If we recall that $\Phi(\sigma) = \Phi(T^*)e^{-\lambda(t-T^*)}$, we find that:

$$\pi(T^*) = \mu\Phi(T^*)/\lambda \ .$$

The process' continuous-time statistical complexity—precisely, entropy of this mixed random variable—is given by:

$$C_\mu = \int_0^{T^*}\mu\Phi(\sigma)\log\frac{1}{\mu\Phi(\sigma)}d\sigma - \frac{\mu\Phi(T^*)}{\lambda}\log\frac{\mu\Phi(T^*)}{\lambda} \ .$$

This is the sum of the nondivergent C_μ component and the rate of divergence of C_μ of the infinitesimal time-discretized renewal process [9].

C. Eventually-Δ Poisson

Probability wave propagation equations, like those in Eq. (6), hold for $\sigma < T^*$ and for $T^* < \sigma < T^* + \Delta$. At $\sigma = T^*$, if no event is observed, probability flows in from both $(T^*+\Delta)^-$ and from $(T^*)^-$, giving rise to the equation:

$$\rho(T^*,t+\Delta t) = \rho(T^*-\Delta t,t) + \rho(T^*+\Delta^*-\Delta t,t) \ .$$

Unfortunately, there is a discontinuous jump in $\rho(\sigma,t)$ at $\sigma = T^*$ coming from $(T^*)^-$ and $(T^*+\Delta^*)^-$. And so, we cannot Taylor expand either $\rho(T^*-\Delta t,t)$ or $\rho(T^*+\Delta^*-\Delta t,t)$ about $\Delta t = 0$.

Again, we can use the prescription suggested by Thm. 1 to calculate the probability density function over these causal states and, from that, calculate the continuous-time statistical complexity. Below $\sigma < T^*$, the probability density function over causal states is exactly that described in Sec. IV A: $\rho(\sigma) = \mu\Phi(\sigma)$. For $T^* \leq \sigma < T^* + \Delta$, the probability density function becomes:

$$\begin{aligned}\rho(\sigma) &= \sum_{\sigma':(\sigma'-T^*)\bmod \Delta^*=\sigma}\mu\Phi(\sigma') \\ &= \mu\sum_{i=0}^{\infty}\Phi(\sigma+i\Delta^*) \ .\end{aligned}$$

Recalling Def. 2, we see that $\Phi(\sigma+i\Delta^*) = e^{-\lambda i}\Phi(\sigma)$ and so find that for $\sigma > T^*$:

$$\begin{aligned}\rho(\sigma) &= \mu\Phi(\sigma)\sum_{i=0}^{\infty}e^{-\lambda i} \\ &= \frac{\mu\Phi(\sigma)}{1-e^{-\lambda}} \ .\end{aligned}$$

Altogether, this gives the statistical complexity:

$$\begin{aligned}C_\mu = &\int_0^{T^*}\mu\Phi(\sigma)\log\frac{1}{\mu\Phi(\sigma)}d\sigma \\ &+ \int_{T^*}^{T^*+\Delta^*}\frac{\mu\Phi(\sigma)}{1-e^{-\lambda}}\log\frac{1-e^{-\lambda}}{\mu\Phi(\sigma)}d\sigma \ .\end{aligned}$$

V. DIFFERENTIAL INFORMATION RATES

We define continuous-time information anatomy [19] quantities as *rates*. As mentioned earlier, the present extends over an infinitesimal time. To define information anatomy rates, we let Γ_δ be the symbols observed over an arbitrarily small length of time δ, starting at the present 0^-. It could be that Γ_δ encompasses some portion of $\mathcal{T}_1$;

Quantity	Expression	
$C_\mu^+ = \mathrm{H}[\mathcal{S}^+]$	$\int_0^\infty \mu\Phi(\sigma)\log\frac{1}{\mu\Phi(\sigma)}d\sigma$	Not eventually Δ-Poisson
	$\int_0^{T^*} \mu\Phi(\sigma)\log\frac{1}{\mu\Phi(\sigma)}d\sigma - \frac{\mu\Phi(T^*)}{\lambda}\log\frac{\mu\Phi(T^*)}{\lambda}$	Eventually Poisson
	$\int_0^{T^*} \mu\Phi(\sigma)\log\frac{1}{\mu\Phi(\sigma)}d\sigma + \int_{T^*}^{T^*+\Delta^*} \frac{\mu\Phi(\sigma)}{1-e^{-\lambda}}\log\frac{1-e^{-\lambda}}{\mu\Phi(\sigma)}d\sigma$	Eventually Δ-Poisson
$\mathbf{E} = \mathrm{I}[\mathcal{T}_{:0^+};\mathcal{T}_{0^-:}]$	$\int_0^\infty \mu t\phi(t)\log_2\big(\mu\phi(t)\big)dt - 2\int_0^\infty \mu\Phi(t)\log_2\big(\mu\Phi(t)\big)dt$	
$h_\mu = \lim_{\delta\to0}\frac{d\,\mathrm{H}[\Gamma_\delta\|\mathcal{T}_{:0^+}]}{d\delta}$	$-\mu\int_0^\infty \phi(t)\log\phi(t)dt$	
$b_\mu = \lim_{\delta\to0}\frac{d\,\mathrm{I}[\mathcal{T}_{\delta^-:};\Gamma_\delta\|\mathcal{T}_{:0^+}]}{d\delta}$	$-\mu\Big(2\int_0^\infty \phi(t)\log\phi(t+t')dt - 1 - \int_0^\infty\phi(t)\int_0^\infty\phi(t')\log\phi(t+t')dt'dt\Big)$	
$q_\mu = \lim_{\delta\to0}\frac{d\,\mathrm{I}[\mathcal{T}_{:0^+};\Gamma_\delta;\mathcal{T}_{\delta^-:}]}{d\delta}$	$\mu\int_0^\infty\phi(t)\int_0^\infty\phi(\sigma^+)\log\phi(t')dt'dt - \mu\log\mu$	
$r_\mu = \lim_{\delta\to0}\frac{d\,\mathrm{H}[\Gamma_\delta\|\mathcal{T}_{:0^+},\mathcal{T}_{\delta^-:}]}{d\delta}$	$-\mu\Big(2\int_0^\infty \phi(t)\log\phi(t+t')dt - 1 - \int_0^\infty\phi(t)\int_0^\infty\phi(t')\log\phi(t+t')dt'dt\Big)$	
$H_0 = \lim_{\delta\to0}\frac{d\,\mathrm{H}[\Gamma_\delta]}{d\delta}$	$\mu - \mu\log\mu$	

TABLE I. Information measures and differential information rates of continuous-time renewal processes. See Sec. V for calculational details. Several information measures are omitted as they are linear combinations of information measures already in the table; e.g., $\rho_\mu = b_\mu + q_\mu$ [19].

the notation leaves this ambiguous. The entropy rate is now:

$$h_\mu = \lim_{\delta\to0}\frac{d\,\mathrm{H}[\Gamma_\delta|\mathcal{T}_{:0^+}]}{d\delta} \ . \qquad (8)$$

This is equivalent to the more typical random-variable "block" definition of entropy rate [11]: $\lim_{\delta\to\infty}\mathrm{H}[\mathcal{T}_{0:\delta}]/\delta$. Similarly, we define the *single-measurement entropy rate* as:

$$\mathrm{H}_0 = \lim_{\delta\to0}\frac{d\,\mathrm{H}[\Gamma_\delta]}{d\delta} \ , \qquad (9)$$

the *bound information rate* as:

$$b_\mu = \lim_{\delta\to0}\frac{d\,\mathrm{I}[\mathcal{T}_{\delta^-:};\Gamma_\delta|\mathcal{T}_{:0^+}]}{d\delta} \ , \qquad (10)$$

the *ephemeral information rate* as:

$$r_\mu = \lim_{\delta\to0}\frac{d\,\mathrm{H}[\Gamma_\delta|\mathcal{T}_{:0^+},\mathcal{T}_{\delta^-:}]}{d\delta} \ , \qquad (11)$$

and the *co-information rate* as:

$$q_\mu = \lim_{\delta\to0}\frac{d\,\mathrm{I}[\mathcal{T}_{:0^+};\Gamma_\delta;\mathcal{T}_{\delta^-:}]}{d\delta} \ . \qquad (12)$$

In direct analogy to discrete-time process information anatomy, we have the relationships:

$$\mathrm{H}_0 = 2b_\mu + r_\mu + q_\mu \ ,$$
$$h_\mu = b_\mu + r_\mu \ .$$

So, the entropy rate h_μ, the instantaneous rate of information creation, again decomposes into a component b_μ that represents active information storage and a component r_μ that represents "wasted" information.

Prescient states (not necessarily *minimal*) are adequate for deriving all information measures aside from $C_\mu^\pm$. As such, we focus on the transition dynamics of non-eventually Δ-Poisson ϵ-machines and, implicitly, their bidirectional machines.

To find the joint probability density function of the time to next event σ^- and time since last event σ^+, we note that $\sigma^+ + \sigma^-$ is an interevent interval; hence:

$$\rho(\sigma^+,\sigma^-) \propto \phi(\sigma^+ + \sigma^-) \ .$$

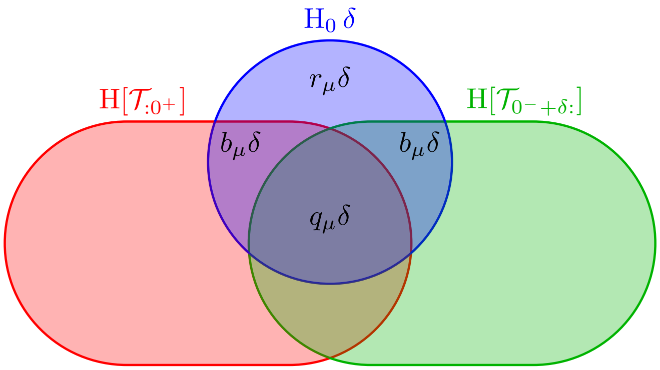

FIG. 6. Predictively useful and predictively useless randomness for renewal processes ($|\mathcal{A}| = 1$): Information diagram for the past $\mathcal{T}_{:0^+}$, infinitesimal present Γ_δ, and future $\mathcal{T}_{\delta^-:}$. The measurement entropy rate H_0 is the rate of change of the single-measurement entropy $H[\Gamma_\delta]$ at $\delta = 0$. The ephemeral information rate $r_\mu = H[\Gamma_\delta|\mathcal{T}_{:0^+}, \mathcal{T}_{\delta^-:}]$ is the rate of change of useless information generation at $\delta = 0$. The bound information rate $b_\mu = I[\Gamma_\delta; \mathcal{T}_{\delta^-:}|\mathcal{T}_{:0^+}]$ is the rate of change of active information storage. And, the co-information rate $q_\mu = I[\mathcal{T}_{:0^+}; \Gamma_\delta; \mathcal{T}_{\delta^-:}]$ is the rate of change of shared information between past, present, and future. These definitions closely parallel those in Ref. [19].

The normalization factor of this distribution is:

$$\begin{aligned} Z &= \int_0^\infty \int_0^\infty \phi(\sigma^+ + \sigma^-) d\sigma^+ d\sigma^- \\ &= \int_0^\infty \int_{\sigma^-}^\infty \phi(\sigma^+) d\sigma^+ d\sigma^- \\ &= \int_0^\infty \sigma^- \phi(\sigma^-) d\sigma^- \\ &= \mu^{-1} \ . \end{aligned}$$

So, the joint probability distribution is:

$$\begin{aligned} \rho(\sigma^+, \sigma^-) &= \frac{\phi(\sigma^+ + \sigma^-)}{Z} \\ &= \mu\phi(\sigma^+ + \sigma^-) \ . \end{aligned}$$

Equivalently, we could have calculated the conditional probability density function of time-to-next-event given that it has been at least σ^+ since the last event. This, by similar arguments, is $\phi(\sigma^+ + \sigma^-)/\Phi(\sigma^+)$. This would have given the same expression for $\rho(\sigma^+, \sigma^-)$.

To find the excess entropy, we merely need calculate [12, 13]:

$$\begin{aligned} \mathbf{E} &= I[\mathcal{S}^+; \mathcal{S}^-] \\ &= \int_0^\infty \int_0^\infty \mu\phi(\sigma^+, \sigma^-) \log \frac{\mu\phi(\sigma^+, \sigma^-)}{\phi(\sigma^+)\phi(\sigma^-)} d\sigma^+ d\sigma^- \ . \end{aligned}$$

Algebra not shown here gives:

$$\begin{aligned} \mathbf{E} = \int_0^\infty & \mu\, t\, \phi(t) \log_2 \big(\mu\, \phi(t)\big) dt \\ & - 2 \int_0^\infty \mu\, \Phi(t) \log_2 \big(\mu\, \Phi(t)\big) dt \ . \end{aligned}$$

Unsurprisingly [20], this agrees with the formula given in Ref. [9], which was derived by considering the limit of infinitesimal time discretization.

Now, we turn to the more technically challenging task of calculating differential information anatomy rates. Suppose that Γ_δ is a random variable for paths of length δ. Each path is uniquely specified by a list of times of events. Let X_δ be a random variable defined by:

$$X_\delta = \begin{cases} 0 & \text{No events in } \Gamma_\delta \\ 1 & \text{1 event in } \Gamma_\delta \\ 2 & \geq 2 \text{ events in } \Gamma_\delta \end{cases} \ .$$

We first illustrate how to find H_0, since the same technique allows calculating h_μ. We can rewrite the path entropy as:

$$H[\Gamma_\delta] = H[X_\delta] + H[\Gamma_\delta|X_\delta] \ .$$

For renewal processes, when μ can be defined, we see that:

$$\begin{aligned} &\Pr(X_\delta = 0) = 1 - \mu\delta + O(\delta^2) \ , \\ &\Pr(X_\delta = 1) = \mu\delta + O(\delta^2) \ , \text{ and} \\ &\Pr(X_\delta = 2) = O(\delta^2) \ . \end{aligned}$$

Straightforward algebra shows that:

$$H[X_\delta] = \mu\delta - \mu\delta\log(\mu\delta) + O(\delta^2 \log\delta) \ .$$

We would like to find a similar asymptotic expansion for $H[\Gamma_\delta|X_\delta]$, which can be rewritten as:

$$\begin{aligned} H[\Gamma_\delta|X_\delta] = \Pr(X_\delta = 0)\, & H[\Gamma_\delta|X_\delta = 0] \\ & + \Pr(X_\delta = 1)\, H[\Gamma_\delta|X_\delta = 1] \\ & + \Pr(X_\delta = 2)\, H[\Gamma_\delta|X_\delta = 2] \ . \end{aligned}$$

First, we notice that Γ_δ is deterministic given that $X_\delta = 0$—the path of all silence. So, $H[\Gamma_\delta|X_\delta = 0] = 0$. Second, we can similarly ignore the term $\Pr(X_\delta = 2)\, H[\Gamma_\delta|X_\delta = 2]$ since $\Pr(X_\delta = 2)$ is $O(\delta^2)$ and, we claim, $H[\Gamma_\delta|X_\delta = 2]$ is $O(\log\delta)$: by standard maximum entropy arguments, $H[\Gamma_\delta|X_\delta = 2]$ is at most $\log\delta$, and by noting that trajectories with only one event are a strict subset of trajectories with more than one event but with multiple events arbitrarily close to one another, $H[\Gamma_\delta|X_\delta = 2] \geq$

$\mathrm{H}[\Gamma_\delta|X_\delta = 1]$ which, by arguments below, is $O(\log\delta)$. Thus, the term $\Pr(X_\delta = 2)\,\mathrm{H}[\Gamma_\delta|X_\delta = 2]$ is $O(\delta^2\log\delta)$ at most. Finally, to calculate $\mathrm{H}[\Gamma_\delta|X_\delta = 1]$, we note that when $X_\delta = 1$, paths can be uniquely specified by an event time, whose probability is $\Pr(\mathcal{T} = t|X_\delta = 1) \propto \Phi(t)\Phi(\delta - t)$. A Taylor expansion about $\delta/2$ shows that $\Pr(\mathcal{T} = t|X_\delta = 1) = \frac{1}{\delta} + h(t)$ for some $h(t)$ in which $\lim_{\delta\to 0}\delta^3 h(t) = 0$ for all $0 \leq t \leq \delta$. So, overall, we find that:

$$\Pr(\Gamma_\delta|X_\delta = 1) = \frac{1}{\delta} + \delta\Delta\Pr(\Gamma_\delta|X_\delta = 1) \ ,$$

where $\lim_{\delta\to 0}\delta^2\Delta\Pr(\Gamma_\delta|X_\delta = 1) = 0$ for any Γ_δ with at least one event in the path. The largest corrections to $\Pr(\Gamma_\delta|X_\delta = 1)$ come from ignoring the paths with two or more events, rather than from approximating all paths with only one event as equally likely. In sum, we see that:

$$\mathrm{H}[\Gamma_\delta|X_\delta] = \mu\delta\log\delta + O(\delta^2\log\delta) \ .$$

Together, these manipulations give:

$$\mathrm{H}[\Gamma_\delta] = \mu\delta - \mu\delta\log\mu + O(\delta^2\log\delta) \ .$$

This then implies:

$$\begin{aligned} H_0 &= \lim_{\delta\to 0}\frac{d\,\mathrm{H}[\Gamma_\delta]}{d\delta} \\ &= \mu - \mu\log\mu \ . \end{aligned}$$

A similar series of arguments helps to calculate $h_\mu(\sigma^+)$ defined in Eq. (8), where now, μ is replaced by $\phi(\sigma^+)/\Phi(\sigma^+)$:

$$h_\mu(\sigma^+) = \frac{\phi(\sigma^+)}{\Phi(\sigma^+)} - \frac{\phi(\sigma^+)}{\Phi(\sigma^+)}\log\frac{\phi(\sigma^+)}{\Phi(\sigma^+)} \ , \qquad (13)$$

which gives:

$$\begin{aligned} h_\mu = &\int_0^\infty \mu\phi(\sigma^+)d\sigma^+ \\ &- \int_0^\infty \mu\phi(\sigma^+)\log\frac{\phi(\sigma^+)}{\Phi(\sigma^+)}d\sigma^+ \ . \end{aligned}$$

Algebra (namely, integration by parts) not shown here yields the expression:

$$h_\mu = -\mu\int_0^\infty \phi(t)\log\phi(t)dt \ . \qquad (14)$$

As expected, this is the nondivergent component of the expression given in Eq. (10) of Ref. [9] for the δ-entropy rate of renewal processes. And, it agrees with expressions derived in alternative ways [21].

We need slightly different techniques to calculate b_μ, as we no longer need to decompose a path entropy. From Eq. (10), we have:

$$b_\mu(\sigma^+) = \lim_{\delta\to 0}\frac{d\,\mathrm{H}[\mathcal{S}^-_\delta|\mathcal{S}^+_0 = \sigma^+]}{d\delta} \ .$$

Let's develop a short-time δ asymptotic expansion for $\Pr(\mathcal{S}^-_\delta = \sigma^-|\mathcal{S}^+_0 = \sigma^+)$. First, we notice that $\mathcal{S}^+_0 \to \mathcal{S}^+_\delta \to \mathcal{S}^-_\delta$, so that:

$$\begin{aligned} &\Pr(\mathcal{S}^-_\delta = \sigma^-|\mathcal{S}^+_0 = \sigma^+) \\ &\quad = \int_0^\infty \Pr(\mathcal{S}^-_\delta = \sigma^-|\mathcal{S}^+_\delta = \sigma')\Pr(\mathcal{S}^+_\delta = \sigma'|\mathcal{S}^+_0 = \sigma^+)d\sigma'. \end{aligned}$$

We already can identify:

$$\Pr(\mathcal{S}^-_\delta = \sigma^-|\mathcal{S}^+_\delta = \sigma') = \frac{\phi(\sigma^- + \sigma')}{\Phi(\sigma')} \ .$$

To understand $\Pr(\mathcal{S}^+_\delta = \sigma'|\mathcal{S}^+_0 = \sigma')$, we expand:

$$\Pr(\mathcal{S}^+_\delta = \sigma'|\mathcal{S}^+_0 = \sigma^+) = \sum_{x=0}^{2}\Pr(\mathcal{S}^+_\delta = \sigma', X_\delta = x|\mathcal{S}^+_0 = \sigma^+) \ .$$

Recall that $\Pr(X_\delta = 2|\mathcal{S}^+_0 = \sigma^+)$ is $O(\delta^2)$, that:

$$\Pr(\mathcal{S}^+_\delta = \sigma', X_\delta = 0|\mathcal{S}^+_0 = \sigma^+) = \frac{\Phi(\sigma')}{\Phi(\sigma^+)}\delta(\sigma' - \delta - \sigma^+) \ ,$$

and that:

$$\begin{aligned} &\Pr(\mathcal{S}^+_\delta = \sigma', X_\delta = 1|\mathcal{S}^+_0 = \sigma^+) \\ &\qquad = \begin{cases} \frac{\phi(\sigma^+ + \delta - \sigma')}{\Phi(\sigma^+)}\Phi(\sigma') & \sigma' \leq \delta \\ 0 & \sigma' > \delta \end{cases} \ . \end{aligned}$$

Then, straightforward algebra not shown gives:

$$\begin{aligned} &\Pr(\mathcal{S}^-_\delta = \sigma^-|\mathcal{S}^+_0 = \sigma^+) \\ &\quad = \frac{\phi(\sigma^+ + \sigma^-)}{\Phi(\sigma^+)} + \frac{\phi'(\sigma^+ + \sigma^-) + \phi(\sigma^-)\phi(\sigma^+)}{\Phi(\sigma^+)}\delta + O(\delta^2) \ . \end{aligned}$$

This can be used to derive:

$$\begin{aligned} b_\mu(\sigma^+) = \frac{\phi(\sigma^+)}{\Phi(\sigma^+)}\Big(&\log\phi(\sigma^+) - 1 \\ &- \int_0^\infty \phi(\sigma^-)\log\phi(\sigma^+ + \sigma^-)d\sigma^-\Big) \ , \end{aligned}$$

in nats. When $\phi(t) = \lambda e^{-\lambda t}$, for instance, $b_\mu(\sigma^+) = 0$ for all σ^+, confirming in a much more complicated calculation that Poisson processes really are memoryless. This allows us to calculate the total b_μ as:

$$\begin{aligned} b_\mu &= \int_0^\infty \mu \Phi(\sigma^+) b_\mu(\sigma^+) d\sigma^+ \\ &= -\mu \Big(1 + \int_0^\infty \int_0^\infty \phi(t)\phi(t') \log \phi(t+t') dt dt' \\ &\qquad - \int_0^\infty \phi(t) \log \phi(t) dt \Big) ~, \end{aligned}$$

in nats. And, from this, we find r_μ using:

$$\begin{aligned} r_\mu &= h_\mu - b_\mu \\ &= -\mu \int_0^\infty \phi(\sigma^+) \log \phi(\sigma^+) d\sigma^+ \\ &\quad + \mu \Big(1 + \int_0^\infty \int_0^\infty \phi(\sigma^+)\phi(\sigma^-) \log \Phi(\sigma^+) d\sigma^+ d\sigma^- \\ &\qquad - \int_0^\infty \phi(\sigma^+) \log \phi(\sigma^+) d\sigma^+ \Big) \\ &= -\mu \Big(2 \int_0^\infty \phi(t) \log \phi(t+t') dt - 1 \\ &\qquad - \int_0^\infty \phi(t) \int_0^\infty \phi(t') \log \phi(t+t') dt' dt \Big) ~. \end{aligned}$$

Continuing, we calculate q_μ from:

$$\begin{aligned} q_\mu &= \mathrm{H}_0 - (h_\mu + b_\mu) \\ &= -\mu \log \mu - \mu \\ &\quad + \mu \Big(\int_0^\infty \phi(t) \int_0^\infty \phi(t') \log \phi(t+t') dt' dt + 1 \Big) \\ &= \mu \int_0^\infty \phi(t) \int_0^\infty \phi(t') \log \phi(t') dt' dt - \mu \log \mu ~. \end{aligned}$$

And, we calculate ρ_μ via:

$$\begin{aligned} \rho_\mu &= \mathrm{H}_0 - h_\mu \\ &= -\mu \log \mu - \mu + \mu \int_0^\infty \phi(\sigma^+) \log \phi(\sigma^+) d\sigma^+ ~. \end{aligned}$$

All these quantities are gathered in Table I, which gives them in bits rather than nats.

VI. UNIFILAR HIDDEN SEMI-MARKOV MODELS

The ϵ-machines of discrete-time, discrete-symbol processes are well understood and, as we now appreciate from Secs. III-V, the predictive equivalence relation defining them readily applies to continuous-time renewal processes. This gives the latter's analogous maximally predictive models: continuous or hybrid discrete-continuous ϵ-machines, when minimal. Here, we introduce a new class of process generators that are unifilar versions of Ref. [22]'s hidden semi-Markov models, but whose dwell time distributions can take any form. (Note that general semi-Markov models are a strict subset.) Roughly speaking, they are stateful renewal processes, but this needs to be clarified. Many of their calculations reduce to those in Secs. III-V. When appropriate, we skip these steps.

We start by introducing the minimal generative models in Fig. 7. Let $\mathcal{G}$ be the set of states in this generative model. Each state $g \in \mathcal{G}$ emits a symbol $x \in \mathcal{A}$ and a dwell time $\tau \sim \phi_{g,x}$ for that symbol, and, based on the state g and emitted symbol x, transitions to a new state g'. We assume that the underlying generative model is *unifilar*: that the new state g is uniquely specified by the prior state g and emitted symbol x. We introduce a perhaps unfamiliar restriction on the labeled transition matrices $\{T_g^{(x)} : x \in \mathcal{A}\}$. Define $\mathrm{supp}(g) := \{x : p(x|g) > 0\}$ and $\mathrm{supp}(g \to g') := \{x : p(g', x|g) > 0\}$. Then, we focus only on generative models for which $\mathrm{supp}(g) \cap \mathrm{supp}(g \to g') = \emptyset$. This simply ensures that there is no uncertainty in when one dwell time finishes and another begins. For example, consider the generator in Fig. 7(bottom): if states B and C were both to emit a 0 in succession, it would be impossible to tease apart when the process switched from state B to state C. The restriction introduces no loss of generality for our purposes.

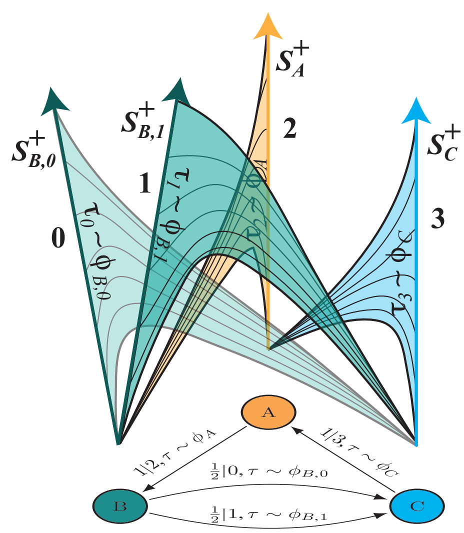

FIG. 7. Unifilar hidden semi-Markov model: (Top) Prescient machine—an ϵ-machine under mild conditions—for the process emitted by the generator below. Causal states $\mathcal{S}^+_A$, $\mathcal{S}^+_{B,0}$, $\mathcal{S}^+_{B,1}$, and $\mathcal{S}^+_C$ are isomorphic to $\Re^+$. During an event interval, symbol $x \in \{0,1,2,3\}$ is emitted. A transition occurs to a new event when the state dwell time is exhausted at τ_x, which is distributed according to $\phi_{\sigma,x}$. (Bottom) Generative model with three hidden states (A, B, and C) emits symbols $x \in \{0,1,2,3\}$ for dwell times τ drawn from probability density functions ϕ_A, $\phi_{B,0}$,$\phi_{B,1}$, and ϕ_C, respectively. (Transition labels as in Fig. 1.)

A prescient model of this combined process is shown in Fig. 7(top). Each state $g \in \mathcal{G}$ comes equipped with one or more renewal process-like tails (semi-infinite spaces that act as continuous counters) that generically take the form of Fig. 3. The leakiness of these (dissipative) counters is given by $\phi_{g,x}$, the probability density function from which the dwell time is drawn. This new form of state-transition diagram depicts the ϵ-machine of these hidden semi-Markov processes. Moreover, if one or more of the dwell-time distributions gives an eventually Poisson or an eventually-Δ Poisson structure, the presentation in Fig. 7(top) is a prescient machine, but not the ϵ-machine.

More generally, any such unifilar minimal generative model has a prescient machine with a "node" for each underlying hidden state g and as many counters as needed—one for every almost-everywhere unique $\phi_{g,\cdot}$. Each counter leaks probability to the next underlying hidden state g', which is completely determined by g and x.

Theorem 2. *The presentation in Fig. 7(top) is a prescient machine for the process generated by the unifilar hidden semi-Markov model of Fig. 7(bottom).*

Proof. *To show that this is a prescient machine, we need to show that the present model state– consisting of hidden state g, current emitted symbol x, and dwell time τ– is uniquely specified by the observed past almost surely. The observed symbol x is given by the current symbol in the observed past. The restriction on successive emitted symbols (that $supp(g) \cap supp(g \to g') = \emptyset$) implies that the observed dwell time τ is exactly the observed length of x. Finally, the underlying hidden state g is determined uniquely by a* function *of the past almost surely, in which all dwell-time information is removed, by assumption: the restriction mentioned earlier implies there is no uncertainty in when one dwell time finishes and another begins. And, the unifilarity of the dynamic on hidden states g implies that the sequence of symbols in the observed past are sufficient to specify the hidden state g almost surely. Hence, g is determined uniquely from the observed past almost surely. The theorem follows.*

Remark. *Theorem 2 can be straightforwardly generalized to specify conditions under which the presentation is an ϵ-machine, a minimal prescient machine, by incorporating the conditions of Thm. 1.*

The stationary distribution for $\rho(\tau|g,x)$ directly follows the treatment for the continuous-time renewal processes in Sec. IV A, and so:

$$\rho(\tau|g,x) = \mu_{g,x}\Phi_{g,x}(\tau) \ ,$$

where $\mu_{g,x} = 1/\int_0^\infty \Phi_{g,x}(\tau)d\tau$. Then we note that:

$$\begin{aligned} p(x|g) &\propto \frac{T_g^{(x)}}{\mu_{g,x}} \\ \to p(x|g) &= \frac{T_g^{(x)}/\mu_{g,x}}{\sum_{x'} T_g^{(x')}/\mu_{g,x'}} \ . \end{aligned}$$

And so:

$$\begin{aligned} \rho(g,x,\tau) &= p(g)\frac{T_g^{(x)}/\mu_{g,x}}{\sum_{x'} T_g^{(x')}/\mu_{g,x'}}\mu_{g,x}\Phi_{g,x}(\tau) \\ &= p(g)\frac{T_g^{(x)}}{\sum_{x'} T_g^{(x')}/\mu_{g,x'}}\Phi_{g,x}(\tau) \ . \end{aligned} \quad (15)$$

To find $p(g)$, we again calculate the probability mass dumped at $\tau = 0$ in terms of $p(g)$:

$$p(g,x,0) = \sum_{g',x'} \int_0^\infty p(g',x',\tau)\delta_{\epsilon(g',x'),g}\frac{\phi_{g',x'}(\tau)}{\Phi_{g',x'}(\tau)}T_g^{(x)}d\tau \ .$$

After a straightforward substitution of Eq. (15) and noting that $\int_0^\infty \phi_{g,x}(\tau) = 1$, we find:

$$p(g)\frac{1}{\sum_{x'} T_g^{(x')}/\mu_{g,x'}} = \sum_{g',x'} \frac{1}{\sum_x T_g^{(x)}/\mu_{g,x}} T_{g',g}^{(x')} p(g') \ .$$

So:

$$\frac{p(g)}{\sum_x T_g^{(x)}/\mu_{g,x}} = \sum_{g'} T_{g',g}\frac{p(g')}{\sum_x T_g^{(x)}/\mu_{g,x}} \ .$$

Let $\pi(g)$ be the stationary distribution for the underlying discrete-state ϵ-machine:

$$\pi := \text{eig}_1(T) \ ,$$

where the eigenvector is normalized such that the sum of its entries is 1. Then:

$$\pi(g) \propto \frac{p(g)}{\sum_x T_g^{(x)}/\mu_{g,x}} \ .$$

Or, rewriting and normalizing, we have:

$$p(g) = \pi(g)\frac{\sum_x T_g^{(x)}/\mu_{g,x}}{\sum_{g',x} \pi(g')T_{g'}^{(x)}/\mu_{g',x}} \ .$$

Altogether, we find that the steady-state distribution is given by:

$$\rho(g,x,\tau) = \left(\pi(g)\frac{\sum_{x'} T_g^{(x')}/\mu_{g,x'}}{\sum_{g',x'}\pi(g')T_{g'}^{(x')}/\mu_{g',x'}}\right) \times \left(\frac{T_g^{(x)}/\mu_{g,x}}{\sum_{x'} T_g^{(x')}/\mu_{g,x'}}\right)\mu_{g,x}\Phi_{g,x}(\tau) = \frac{\pi(g)T_g^{(x)}\Phi_{g,x}(\tau)}{\sum_{g',x'}\pi(g')T_{g'}^{(x')}/\mu_{g',x'}} ~. \quad (16)$$

Using the formulae for entropies of mixed random variables [23], we find a statistical complexity of:

$$\begin{aligned} C_\mu^+ &= \mathrm{H}[\rho(g,x,\tau)] \\ &= \langle \mathrm{H}[\rho(\tau|g,x)]\rangle_{p(g,x)} + \mathrm{H}[p(g,x)] \\ &= \left\langle \int_0^\infty \mu_{g,x}\Phi_{g,x}(\tau)\log\frac{1}{\mu_{g,x}\Phi_{g,x}(\tau)}d\tau \right\rangle_{p(g,x)} \\ &\quad + \mathrm{H}\left[\frac{\pi(g)T_g^{(x)}/\mu_{g,x}}{\sum_{g',x'}\pi(g')T_{g'}^{(x')}/\mu_{g',x'}}\right] ~. \end{aligned}$$

Note that $\mathrm{H}[\pi(g)]$ is the statistical complexity of the underlying discrete-time ϵ-machine and that $\mathrm{H}[\rho(\tau|g,x)]$ is the statistical complexity of a noneventually Δ-Poisson renewal process with interevent distribution $\phi_{g,x}(\tau)$, averaged over g and x. Hence, the statistical complexity of these unifilar hidden semi-Markov processes differs from the statistical complexity of its "components" by:

$$\mathrm{H}\left[\frac{\pi(g)T_g^{(x)}/\mu_{g,x}}{\sum_{g',x'}\pi(g')T_{g'}^{(x')}/\mu_{g',x'}}\right] - \mathrm{H}[\pi(g)] ~.$$

Whether this difference is positive or negative depends on both matrices $T_g^{(x)}/\mu_{g,x}$. In general, we expect the difference to be positive.

Since there are multiple observed symbols $x \in \mathcal{A}$ generated by these machines, H_0 (and so ρ_μ) and b_μ as defined in Sec. V diverge. However, the entropy rate h_μ and excess entropy $\mathbf{E}$ as defined in Sec. V do not diverge for processes generated by this restricted class of unifilar hidden semi-Markov models. From the steady-state distribution given in Eq. (16) and from the entropy rate expressions in Eqs. (13)-(14) of Sec. V, we immediately have the entropy rate for these unifilar hidden semi-Markov models:

$$\begin{aligned} h_\mu &= \lim_{\delta\to 0}\frac{d\,\mathrm{H}[\Gamma_\delta|\mathcal{S}_{0^+}^+ = (g,x,\tau)]}{d\delta} \\ &= \sum_{g,x}\rho(g,x)\left(-\mu_{g,x}\int_0^\infty \phi_{g,x}(\tau)\log\phi_{g,x}(\tau)d\tau\right) \\ &= -\frac{\sum_{g,x}\pi(g)T_g^{(x)}\int_0^\infty \phi_{g,x}(\tau)\log\phi_{g,x}(\tau)d\tau}{\sum_{g',x'}\pi(g')T_{g'}^{(x')}/\mu_{g',x'}} ~. \quad (17) \end{aligned}$$

To ground intuition, recall that each state in the underlying ϵ-machine for semi-Markov processes corresponds to a unique observation symbol. Hence, setting π to $\mathrm{eig}_1(T_{g,g'})$ and noting that each g is uniquely associated to some x in Eq. (17) recovers the results of Ref. [24] for the entropy rate of semi-Markov processes, though the notation differs somewhat [25].

The process' excess entropy $\mathbf{E} = \mathrm{I}[\mathcal{S}^+;\mathcal{S}^-]$ can be calculated if we can find the joint probability distribution $\Pr(\sigma^+,\sigma^-)$ of forward- and reverse-time causal states. To this end, we add an additional restriction on the generative model: we focus only on generative models for which $\mathrm{supp}(g') \cap \mathrm{supp}(g\to g') = \emptyset$. With this restriction on labeled transition matrices, the time-reversed ϵ-machine of the process has the same form as the ϵ-machine of the forward-time process, but with a different $\mathcal{G}$. The latter is related to the forward-time $\mathcal{G}$ via manipulations described in Ref. [12]. As such, we can write down $p(\sigma^+,\sigma^-)$:

$$\begin{aligned} &\Pr(\sigma^+ = (\tau,g,x)|\sigma^- = (g',x',\tau')) \\ &\qquad = \Pr(\tau|g,x,\tau')\delta_{x,x'}\Pr(g|g',x',x,\tau') \\ &\qquad = \phi_{g,x}(\tau+\tau')\Pr(g|g',x)\delta_{x,x'} ~, \end{aligned}$$

where we obtain $p(g|g',x)$ from standard methods [12, 26] applied to (only) the dynamic on $\mathcal{G}$. Note that $p(g|g',x',x,\tau')$ reduces to $p(g|g',x')$ as g' and x' uniquely specify the distribution from which τ' is drawn and since $x = x'$. We leave the the final steps to $\mathbf{E}$ as an exercise.

VII. CONCLUSIONS

Though the definition of continuous-time causal states parallels that for discrete-time causal states, continuous-time ϵ-machines and information measures are markedly different from their discrete-time counterparts. Similar technical difficulties arise more generally when describing minimal maximally predictive models of other continuous-time, discrete-symbol processes that are not the continuous-time Markov processes analyzed in Ref. [6]. The resulting ϵ-machines do not appear like conventional HMMs—recall Figs. 3-5 and, especially, Fig. 7(top)—and most of the information measures—excepting the excess entropy—are reinterpreted as differential information rates.

Moreover, the ϵ-machine continuous-time machinery gave us a new way to calculate these information measures. Traditionally, expressions for such information measures come from calculating the time-normalized path entropy of arbitrarily long trajectories; e.g., as in Ref. [24]. Instead, we calculated the path entropy of arbitrarily short trajectories, conditioned on the past. This allowed us to extend the results of Ref. [24] for the entropy rate of continuous-time discrete-output processes to a previously untouched class of processes—unifilar hidden semi-Markov processes.

There are two immediate practical benefits to an in-depth look at the ϵ-machines of continuous-time hidden semi-Markov processes. First, statistical model selection when searching through unifilar hidden Markov models is significantly easier than when searching through nonunifilar Hidden Markov models [27], and these benefits should carry over to the case of continuous-time ϵ-machines. Second, the formulae in Table I and those in Sec. VI provide new approaches to binless plug-in information measure estimation; e.g., following Ref. [28].

The machinery required to use continuous-time ϵ-machines is significantly different than that accompanying the study of discrete-time ϵ-machines. Our results here pave the way toward understanding the difficulties that lie ahead when studying the structure and information in continuous-time processes.

ACKNOWLEDGMENTS

The authors thank the Santa Fe Institute for its hospitality during visits. JPC is an SFI External Faculty member. This material is based upon work supported by, or in part by, the U. S. Army Research Laboratory and the U. S. Army Research Office under contract number W911NF-13-1-0390. SM was funded by a National Science Foundation Graduate Student Research Fellowship, a U.C. Berkeley Chancellor's Fellowship, and the MIT Physics of Living Systems Fellowship.

[1] C. R. Shalizi and J. P. Crutchfield. Computational mechanics: Pattern and prediction, structure and simplicity. *J. Stat. Phys.*, 104:817–879, 2001.

[2] W. Lohr. Properties of the statistical complexity functional and partially deterministic hmms. *Entropy*, 11(3):385–401, 2009.

[3] J. P. Crutchfield, P. Riechers, and C. J. Ellison. Exact complexity: Spectral decomposition of intrinsic computation. *Phys. Lett. A*, 380(9-10):998–1002, 2016.

[4] D. Kelly, M. Dillingham, A. Hudson, and K. Wiesner. A new method for inferring hidden Markov models from noisy time sequences. *PLoS One*, 7(1):e29703, 01 2012.

[5] S. Marzen and J. P. Crutchfield. Information anatomy of stochastic equilibria. *Entropy*, 16:4713–4748, 2014.

[6] P. M. Riechers and J. P. Crutchfield. Beyond the spectral theorem: Decomposing arbitrary functions of nondiagonalizable operators. 2016. arxiv.org:1607.06526 [math-ph].

[7] S. Marzen and J. P. Crutchfield. Informational and causal architecture of discrete-time renewal processes. *Entropy*, 17(7):4891–4917, 2015.

[8] S. E. Marzen and J. P. Crutchfield. Statistical signatures of structural organization: The case of long memory in renewal processes. *Phys. Lett. A*, 380(17):1517–1525, 2016.

[9] S. Marzen, M. R. DeWeese, and J. P. Crutchfield. Time resolution dependence of information measures for spiking neurons: Scaling and universality. *Front. Comput. Neurosci.*, 9:109, 2015.

[10] V. Girardin and N. Limnios. Entropy for semi-Markov processes with Borel state spaces: Asymptotic equirepartition properties and invariance principles. *Bernoulli*, 12(3):515–533, 2006.

[11] J. P. Crutchfield and D. P. Feldman. Regularities unseen, randomness observed: Levels of entropy convergence. *CHAOS*, 13(1):25–54, 2003.

[12] C. J. Ellison, J. R. Mahoney, and J. P. Crutchfield. Prediction, retrodiction, and the amount of information stored in the present. *J. Stat. Phys.*, 136(6):1005–1034, 2009.

[13] J. P. Crutchfield, C. J. Ellison, and J. R. Mahoney. Time's barbed arrow: Irreversibility, crypticity, and stored information. *Phys. Rev. Lett.*, 103(9):094101, 2009.

[14] P. Gaspard and X.-J. Wang. Noise, chaos, and (ϵ, τ)-entropy per unit time. *Physics Reports*, 235(6):291–343, 1993.

[15] T. M. Cover and J. A. Thomas. *Elements of Information Theory*. Wiley-Interscience, New York, 1991.

[16] M. Rao, Y. Chen, B. C. Vemuri, and F. Wang. Cumulative residual entropy: a new measure of information. *IEEE Trans. Info. Th.*, 50(6):1220–1228, 2004.

[17] S. B. Lowen and M. C. Teich. Fractal renewal processes generate 1/f noise. *Phys. Rev. E*, 47(2):992–1001, 1993.

[18] H. Risken. *The Fokker-Planck Equation: Methods of Solution and Applications*, volume 18 of *Springer Series in Synergetics*. Springer, 2012.

[19] R. G. James, C. J. Ellison, and J. P. Crutchfield. Anatomy of a bit: Information in a time series observation. *CHAOS*, 21(3):037109, 2011.

[20] M. S. Pinsker. *Information and information stability of random variables and processes*. Holden-Day series in time series analysis. Holden-Day, 1964.

[21] V. Girardin. On the different extensions of the ergodic theorem of information theory. In R. Baeza-Yates, J. Glaz, H. Gzyl, J. Husler, and J. L. Palacios, editors, *Recent Advances in Applied Probability Theory*, pages 163–179. Springer US, 2005.

[22] S. E. Levinson. Continuously variable duration hidden Markov models for automatic speech recognition. *Comp. Speech Lang.*, 1(1):29–45, 1986.

[23] C. Nair, B. Prabhakar, and D. Shah. On entropy for mixtures of discrete and continuous variables. *arXiv:cs/0607075*, 2006.

[24] V. Girardin and N. Limnios. On the entropy for semi-Markov processes. *J. Appl. Prob.*, 40(4):1060–1068, 2003.

[25] Actually, we would apply Eq. (17) to a semi-Markov process in reverse-time so that the underlying model is unifilar rather than co-unifilar; but entropy rate is invariant to time reversal [11].

[26] C. J. Ellison, J. R. Mahoney, R. G. James, J. P. Crutchfield, and J. Reichardt. Information symmetries in irreversible processes. *CHAOS*, 21(3):037107, 2011.

[27] C. C. Strelioff and J. P. Crutchfield. Bayesian structural inference for hidden processes. *Phys. Rev. E*, 89:042119, 2014.

[28] J. D. Victor. Binless strategies for estimation of information from neural data. *Phys. Rev. E*, 66(5):051903, 2002.