

# Informational and Causal Architecture of Continuous-time Renewal Processes

Sarah Marzen<sup>1,2</sup> · James P. Crutchfield<sup>3</sup>

Received: 3 November 2016 / Accepted: 11 April 2017 / Published online: 22 April 2017  
© Springer Science+Business Media New York 2017

**Abstract** We introduce the minimal maximally predictive models ( $\epsilon$ -machines) of processes generated by certain hidden semi-Markov models. Their causal states are either discrete, mixed, or continuous random variables and causal-state transitions are described by partial differential equations. As an application, we present a complete analysis of the  $\epsilon$ -machines of continuous-time renewal processes. This leads to closed-form expressions for their entropy rate, statistical complexity, excess entropy, and differential information anatomy rates.

**Keywords** Renewal process · Entropy rate · Excess entropy · Statistical complexity · Information anatomy · Hidden Semi-Markov models

## 1 Introduction

We are interested in answering two very basic questions about continuous-time renewal processes:

- What are their minimal maximally predictive models—their  $\epsilon$ -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,

---

✉ James P. Crutchfield  
chaos@ucdavis.edu

Sarah Marzen  
semarzen@mit.edu

<sup>1</sup> Physics of Living Systems Group, Department of Physics, Massachusetts Institute of Technology, Cambridge, MA 02139, USA

<sup>2</sup> Department of Physics, University of California at Berkeley, Berkeley, CA 94720, USA

<sup>3</sup> Complexity Sciences Center and Department of Physics, University of California at Davis, One Shields Avenue, Davis, CA 95616, USA

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 seemingly unwieldy case of uncountably infinite causal states.

We analyze continuous-time renewal processes in-depth, as addressing the challenges there carries over to other continuous-time processes. When analyzing discrete-time renewal processes, we can use the  $\epsilon$ -machine definitions outlined in Ref. [1] and the information anatomy definitions outlined in Ref. [7], but neither definitions carry over to the continuous-time case. The difficulties are both technical and conceptual. First, the causal states are now continuous or mixed 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.

Our main thesis is that minimal maximally predictive models of continuous-time renewal processes require a wholly new  $\epsilon$ -machine calculus. To develop it, Sect. 2 describes the required notation and definitions that enable extending computational mechanics, which is otherwise well understood for discrete-time processes [1, 8]. Sections 3, 4, and 5 determine the causal and informational architecture of continuous-time renewal processes. We conclude by describing the challenges overcome and benefits to understanding the information measures, using the new formulae of Table 1.

## 2 Background and Notation

A point process is labeled only with times between events:  $\dots, \tau_{-1}, \tau_0, \tau_1, \dots$ . We view the time series  $\overleftarrow{\tau}$  as a realization of random variables  $\overleftarrow{\mathcal{T}} = \dots, \mathcal{T}_{-1}, \mathcal{T}_0, \mathcal{T}_1, \dots$ . When the observed time series is strictly stationary and the process ergodic, we can calculate the probability distribution  $\Pr(\overleftarrow{\mathcal{T}})$  from a single realization  $\overleftarrow{\tau}$ .

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  $\mathcal{T}_{-\infty:0+} = \dots, \mathcal{T}_{-1}, \mathcal{T}_{0+}$  as the *past* and  $\mathcal{T}_{0-:\infty} = \mathcal{T}_{0-}, \mathcal{T}_1, \dots$  as the *future*. (To reduce notation, we drop the  $\infty$  indices.) The *present*  $\mathcal{T}_{0+:0-}$  itself extends over an infinitesimally small length of time. See Fig. 1.

Continuous-time renewal processes, a special kind of point process, 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*  $\mu$  as:

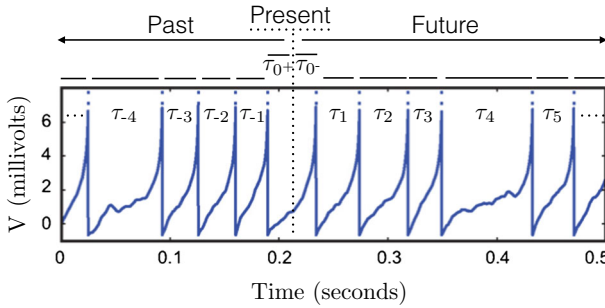
$$\mu^{-1} = \int_0^\infty t\phi(t)dt .$$

Finally, we briefly recall the definitions of entropy, conditional entropy, and mutual information. The entropy of a discrete random variable  $X$  with probability distribution  $p(x)$  is  $H[X] = -\sum_x p(x) \log p(x)$ ; the entropy of a continuous random variable  $X$  with probability density function  $\rho(x)$  is  $H[X] = -\int \rho(x) \log \rho(x)dx$ ; and the entropy of a mixed random variable  $X$  with “probability density function”  $\rho(x)$  and “probability distribution”  $p(x)$  with  $\int \rho(x)dx + \sum p(x) = 1$  was defined in Ref. [9] as  $H[X] =$

**Table 1** Information measures and differential information rates of continuous-time renewal processes

Quantity	Expression	
$C_{\mu}^+ = H[S^+]$	$\int_0^{\infty} \mu \Phi(\sigma) \log \frac{1}{\mu \Phi(\sigma)} d\sigma$	Not eventually $\Delta$ -Poisson
$h_{\mu} = \lim_{\delta \rightarrow 0} \frac{dH[\Gamma_{\delta}   \mathcal{T}_{0^+}]}{d\delta}$	$\int_0^{T^*} \mu \Phi(\sigma) \log \frac{1}{\mu \Phi(\sigma)} d\sigma - \frac{\mu \Phi(T^*)}{\lambda} \log \frac{\mu \Phi(T^*)}{\lambda}$	Eventually Poisson
$b_{\mu} = \lim_{\delta \rightarrow 0} \frac{dH[\mathcal{T}_{\delta}; \Gamma_{\delta}   \mathcal{T}_{0^+}]}{d\delta}$	$\int_0^{T^*} \mu \Phi(\sigma) \log \frac{1}{\mu \Phi(\sigma)} d\sigma + \int_{T^*}^{T^* + \Delta^*} \frac{\mu \Phi(\sigma)}{1 - e^{-\lambda}} \log \frac{1 - e^{-\lambda}}{\mu \Phi(\sigma)} d\sigma$	Eventually $\Delta$ -Poisson
$q_{\mu} = \lim_{\delta \rightarrow 0} \frac{dH[\mathcal{T}_{0^+}; \Gamma_{\delta}; \mathcal{T}_{\delta; \cdot}]}{d\delta}$	$\int_0^{\infty} \mu \phi(t) \log_2 (\mu \phi(t)) dt - 2 \int_0^{\infty} \mu \Phi(t) \log_2 (\mu \Phi(t)) dt$	
$r_{\mu} = \lim_{\delta \rightarrow 0} \frac{dH[\Gamma_{\delta}   \mathcal{T}_{0^+}, \mathcal{T}_{\delta; \cdot}]}{d\delta}$	$-\mu \int_0^{\infty} \phi(t) \log \phi(t) dt$	
$H_0 = \lim_{\delta \rightarrow 0} \frac{dH[\Gamma_{\delta}]}{d\delta}$	$-\mu \left( 2 \int_0^{\infty} \phi(t) \log \phi(t + t') dt - 1 - \int_0^{\infty} \phi(t) \int_0^{\infty} \phi(t') \log \phi(t + t') dt' dt \right)$	
$\rho_{\mu} = \lim_{\delta \rightarrow 0} \frac{dH[\Gamma_{\delta}; \mathcal{T}_{0^+}]}{d\delta}$	$\mu \int_0^{\infty} \phi(t) \int_0^{\infty} \phi(\sigma^+) \log \phi(t') dt' dt - \mu \log \mu$	
	$-\mu \left( 2 \int_0^{\infty} \phi(t) \log \phi(t + t') dt - 1 - \int_0^{\infty} \phi(t) \int_0^{\infty} \phi(t') \log \phi(t + t') dt' dt \right)$	
	$\mu - \mu \log \mu$	
	$-\mu \log \mu - \mu + \mu \int_0^{\infty} \phi(\sigma^+) \log \phi(\sigma^+) d\sigma +$	

See Sect. 5 for calculational details. Several information measures are omitted as they are linear combinations of information measures already in the table; e.g.,  $\sigma_{\mu} = \mathbf{E} - b_{\mu} - q_{\mu}$  [20]



**Fig. 1** Interevent intervals  $\tau_i$  in the past and future and how they relate to the present; shown as they would be identified in a neural spike train (blue line). The  $0^h$  interval, of total length  $\tau_0 = \tau_{0+} + \tau_{0-}$ , is split by the present marker into a time since last event  $\tau_{0+}$  and a time until next event  $\tau_{0-}$  (Color figure online)



**Fig. 2** The smallest generative model of a continuous-time renewal process consists of a single causal state. The transition is labeled  $p|s, \tau \sim \Pr(T)$ , denoting the transition is taken with probability  $p$ , emits symbol  $s$  for duration  $\tau$  distributed according to  $\Pr(T)$ . The length  $T_i = \tau$  of periods of silence (corresponding to output symbol 0) are drawn independently, identically distributed (IID) from probability density  $\phi(t)$

$-\sum p(x) \log p(x) - \int \rho(x) \log \rho(x) dx$ . Conditional entropy of a random variable  $X$  with respect to random variable  $Y$  is, as above,  $H[X|Y] = \langle H[X|Y = y] \rangle_y$ . Mutual information between random variable  $X$  and random variable  $Y$  is  $\mathbf{I}[X; Y] = H[X] - H[X|Y]$  or, equivalently,  $\mathbf{I}[X; Y] = H[Y] - H[Y|X]$ .

**2.1 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:

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

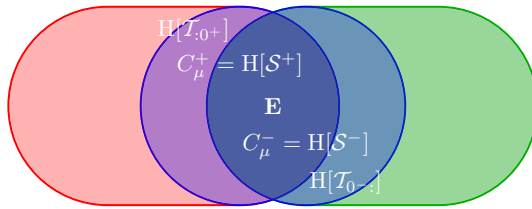
This partitions the set of allowed pasts. Each equivalence class of pasts is a forward-time causal state  $\sigma^+ = \epsilon^+(\tau_{:0+})$ , in which  $\epsilon^+(\cdot)$  is the function that maps a past to its causal state. The set of forward-time causal states  $\mathcal{S}^+ = \{\sigma^+\}$  inherits a probability distribution  $\Pr(\mathcal{S}^+)$  from the probability distribution over pasts  $\Pr(\mathcal{T}_{:0+})$ .

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:

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

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.

The smallest 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. 2. Moreover, the forward- and reverse-time causal states are the same.



**Fig. 3** Predictability, compressibility, and causal irreversibility in renewal processes graphically illustrated using a Venn-like *information diagram* over the random variables for the past  $\mathcal{T}_{0+}$  (left oval, red), the future  $\mathcal{T}_{0-}$  (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. [10].) 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  $\mathbf{E} = \mathbf{I}[\mathcal{T}_{0+}; \mathcal{T}_{0-}]$  is a measure of process predictability (central pointed ellipse, dark blue) and Theorem 1 of Ref. [10, 11] shows that  $\mathbf{E} = \mathbf{I}[\mathcal{S}^+; \mathcal{S}^-]$  by applying the causal shielding relations in Eqs. (1) and (2) (Color figure online)

*Forward-time prescient statistics* are any refinement of the forward-time causal-state partition. By construction, they are a sufficient statistic for prediction, but are not necessarily *minimal* sufficient statistics [1]. *Reverse-time prescient statistics* are any refinement of the reverse-time causal-state partition. They are sufficient statistics for retrodiction, but are again not necessarily minimal.

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

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

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

The consequence of these is illustrated in Fig. 3. Causal shielding holds not just for forward- and reverse-time causal states, but for forward- and reverse-time prescient statistics as well. However, these causal shielding relations are special to prescient statistics, causal states, and their defining functions  $\epsilon^+(\cdot)$  and  $\epsilon^-(\cdot)$ . That is, arbitrary functions of the past and future do not shield the two aggregate past and future random variables from one another. Forward- and reverse-time generative models do not, in general, have state spaces that satisfy Eqs. (1) and (2).

The *forward-time  $\epsilon$ -machine* is that with state space  $\mathcal{S}^+$  and transition dynamic between forward-time causal states. The *reverse-time  $\epsilon$ -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 Sects. 3, 4, and 5.

### 2.2 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. 3 and 7. Many lose meaning when naively applied to continuous-time processes; e.g., see Refs. [5, 12, 13]. We redefine many of these in order to avoid trivial divergences and zeros in Sect. 5.

The *forward-time statistical complexity*  $C_{\mu}^+ = H[\mathcal{S}^+]$  is the cost of coding the forward-time causal states and the *reverse-time statistical complexity*  $C_{\mu}^- = H[\mathcal{S}^-]$  is the cost of coding reverse-time causal states. When  $\mathcal{S}^+$  or  $\mathcal{S}^-$  are mixed or continuous random vari-

ables, one employs differential entropies for  $H[\cdot]$ . The result, though, is that the statistical complexities are potentially negative or infinite or both [14, Chap. 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. [15] for possible alternatives for  $H[\cdot]$ .

### 3 Continuous-time Causal States

Discrete-time renewal processes are temporally symmetric [16], and the same is true for continuous-time renewal processes. As such, we will refer to forward-time causal states and the forward-time  $\epsilon$ -machine as simply causal states or the  $\epsilon$ -machine, with the understanding that reverse-time causal states and reverse-time  $\epsilon$ -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. [16]. 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:

$$\Pr(\mathcal{T}_{0-}|\mathcal{T}_{0+}) = \Pr(\mathcal{T}_1) \Pr(\mathcal{T}_{0-}|\mathcal{T}_{0+}) . \tag{3}$$

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.  $\square$

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 [18].

Intermediate between these two extremes is a broad spectrum 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. [16]’s classification.

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

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

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

Note that a Poisson process is an eventually Poisson renewal process with  $T = 0$ . Another perhaps familiar, but degenerate example of an eventually Poisson renewal process is found in the spike trains emitted by Poisson neurons with refractory periods [13]. There, the neuron is effectively prevented from firing two spikes within a time  $T$  of each other—the period during which 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 and so Poisson. 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 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. Moreover, the renewal process that operates before  $T$  (during the refractory period) is degenerate, allowing no spiking.

Eventually  $\Delta$ -Poisson processes described in Definition 2 are far less intuitive, as  $\phi(t)$  is discontinuous.

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

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

almost everywhere, for the smallest possible  $T^*$  for which  $\Delta^*$  exists.

Finally, we categorize all other renewal processes as “typical” in Definition 3. For instance, the process generated by any generalized integrate-and-fire neuron is typical.

**Definition 3** A typical renewal process is neither eventually Poisson nor eventually  $\Delta$ -Poisson.

Theorem 1 shows that Definitions 1, 2, and 3 offer a complete predictive classification of continuous-time renewal processes.

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

1. When the renewal process is typical, 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  $\Delta$ -Poisson, the causal states are the time since last event up until time  $T^*$  and are the times since  $T^* \bmod \Delta$  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 Theorem 1 of Ref. [16]. 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:

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

This immediately implies the theorem’s first case. If a renewal process is *not* eventually  $\Delta$ -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  $(T, \Delta)$  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:

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

and define the parameters  $T^*$  and  $\Delta^*$  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  $\Delta^*$  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  $\Delta$ -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  $\Delta \rightarrow 0$  and we find that:

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

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

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



### 4 Wave Propagation on Continuous-time $\epsilon$ -machines

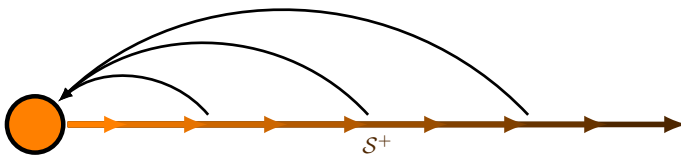
Identifying causal states in continuous-time follows an almost entirely similar path to that used for discrete-time renewal processes in Ref. [16]. The seemingly slight differences between the causal states of eventually Poisson, eventually  $\Delta$ -Poisson, and typical renewal processes, however, have surprisingly important consequences for continuous-time  $\epsilon$ -machines.

As described by Theorem 1, there is often an uncountable infinity of continuous-time causal states. As one might anticipate from Refs. [13, 16], however, there is an ordering to this infinity of causal states that makes calculations tractable. There is one major difference between discrete-time  $\epsilon$ -machines and continuous-time  $\epsilon$ -machines: transition dynamics often amount to specifying the evolution of a probability density function over causal-state space.

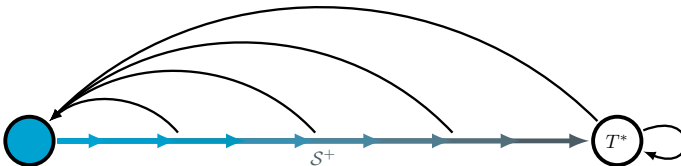
As such, a continuous-time  $\epsilon$ -machine constitutes an unusual presentation of the process generated by a hidden Markov model. It appears as a conveyor belt that transports the distribution of times since last event. Under special conditions, the conveyor belt ends in a trash bin or a second mini-conveyor belt. Compare Figs. 4, 5, and 6.

The exception to this general rule is given by the Poisson process itself. The  $\epsilon$ -machine of a Poisson process is exactly the minimal generative model shown in Fig. 2. 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  $\lambda$ . Hence, the Poisson  $\epsilon$ -machine has only a single state.

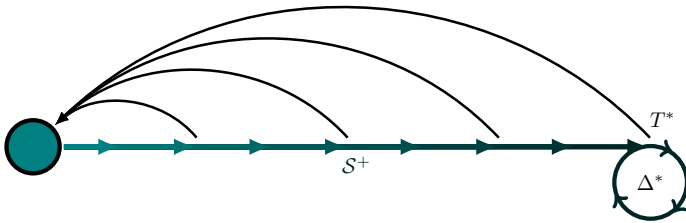
In the general setting, though, the  $\epsilon$ -machine dynamic describes the evolution of the probability density function over its causal states. We therefore search for labeled transition



**Fig. 4**  $\epsilon$ -Machine for the generic not eventually Poisson renewal process: Continuous-time causal states  $S^+$ , 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 orange circle at left*), denoting that an event occurred (Color figure online)



**Fig. 5**  $\epsilon$ -Machine for an eventually Poisson renewal process: Continuous-time causal states  $S^+$  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 blue circle at left*) (Color figure online)



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

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.

Finally, we provide expressions for the forward- and reverse-time statistical complexities of continuous-time renewal processes. Interestingly, a renewal process viewed in reverse-time has equivalent statistics. Therefore, the forward- and reverse-time  $\epsilon$ -machines are equivalent, and so  $C_\mu^+ = C_\mu^-$ . As such, we can relabel both as  $C_\mu$  without fear of confusion.

### 4.1 Typical Renewal Processes

The  $\epsilon$ -machine of a renewal process that is not eventually Poisson takes the state-transition form shown in Fig. 4. Let  $\rho(\sigma, t)$  be the probability density function over the causal states  $\sigma$  at time  $t$ . Our approach to deriving labeled transition dynamics parallels well-known approaches to determining Fokker–Planck equations using a Kramers–Moyal expansion [19]. Here, this means that any probability at causal state  $\sigma$  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:

$$\rho(\sigma, t + \Delta t) = \Pr(S_{t+\Delta t} = \sigma | S_t = \sigma - \Delta t)\rho(\sigma - \Delta t, t). \tag{4}$$

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

$$\Pr(S_{t+\Delta t} = \sigma | S_t = \sigma - \Delta t) = \frac{\Phi(\sigma)}{\Phi(\sigma - \Delta t)}. \tag{5}$$

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:

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

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:

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

Recall that for renewal processes,  $H[S^+] = H[S^-] = C_\mu$ . This was the nondivergent component of the infinitesimal time-discretized renewal process’ statistical complexity found in Ref. [13].

### 4.2 Eventually Poisson Processes

As Theorem 1 anticipates, there is a qualitatively different topology to the  $\epsilon$ -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 Sect. 4.1. 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  $\Delta t$ . 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:

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

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  $t$ , 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  $\lim_{\sigma \rightarrow T^*} \rho(\sigma, t)$ . 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 Theorem 1 to calculate  $\pi(T^*)$  via integration of the stationary distribution over the prescient machine given in Sect. 4.1:

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

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  $C_\mu$  component and the rate of divergence of  $C_\mu$  of the infinitesimal time-discretized renewal process [13].

### 4.3 Eventually- $\Delta$ Poisson Processes

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 Theorem 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 Sect. 4.1:  $\rho(\sigma) = \mu \Phi(\sigma)$ . 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 Definition 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, using Ref. [9], this gives the statistical complexity:

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

### 5 Differential Information Rates

Discrete-time entropy rates are usually defined via  $\lim_{T \rightarrow \infty} H[X_{0:T}]/T$ , which is equivalent to  $H[X_0|X_{:0}]$  by the Cesaro Mean Theorem. The definition for the entropy rate  $h_\mu$  of continuous-time processes follows an analogous pattern. We let  $\Gamma_\delta$  be the symbol sequence observed over an arbitrarily small length of time  $\delta$ , starting at the present  $0^-$ . We have  $h_\mu = \lim_{T \rightarrow \infty} H[\mathcal{T}_{0:T}]/T$ , and the Cesaro Mean Theorem again yields that  $h_\mu = \lim_{\delta \rightarrow 0} dH[\Gamma_\delta | \overleftarrow{\mathcal{T}}]/d\delta$ .

We must similarly redefine other information measures listed in Ref. [20] as differential information rates, where we craft new definitions largely based on Fig. 7. Following Fig. 7 too closely would yield infinities in the final answer that are regularizable; e.g., as described in Ref. [13]. These infinities are exactly related to a well-known problem with differential entropy: the entropy of a coarse-grained continuous random variable  $X$  with probability density function  $\rho(x)$  takes the form of  $-\int \rho(x) \log \rho(x) dx + \log \frac{1}{\epsilon}$ , where  $\epsilon$  is the length of the boxes used in coarse-graining.

Correspondingly, the original set-theoretic definitions [21] do not quite extend to the continuous-time case. As mentioned earlier, the present extends over an infinitesimal time. To define information anatomy rates more precisely, we let  $\Gamma_\delta$  be the symbol sequence observed over an arbitrarily small length of time  $\delta$ , starting at the present  $0^-$ . It could be that  $\Gamma_\delta$  encompasses some portion of  $\mathcal{T}_1$ ; the notation leaves this ambiguous. The entropy rate is now:

$$h_\mu = \lim_{\delta \rightarrow 0} \frac{dH[\Gamma_\delta | \mathcal{T}_{:0+}]}{d\delta}. \tag{8}$$

Again, this is equivalent to the more typical random-variable ‘‘block’’ definition of entropy rate [8]:  $\lim_{T \rightarrow \infty} H[\mathcal{T}_{0:T}]/T$ .

Similarly, we define the *single-measurement entropy rate* as:

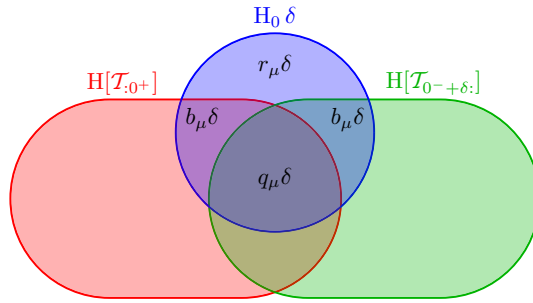
$$H_0 = \lim_{\delta \rightarrow 0} \frac{dH[\Gamma_\delta]}{d\delta}, \tag{9}$$

the *bound information rate* as:

$$b_\mu = \lim_{\delta \rightarrow 0} \frac{d\mathbf{I}[\mathcal{T}_\delta; \Gamma_\delta | \mathcal{T}_{:0+}]}{d\delta}, \tag{10}$$

the *ephemeral information rate* as:

$$r_\mu = \lim_{\delta \rightarrow 0} \frac{dH[\Gamma_\delta | \mathcal{T}_{:0+}, \mathcal{T}_\delta]}{d\delta}, \tag{11}$$



**Fig. 7** Predictively useful and predictively useless information for renewal processes: Information diagram for the past  $\mathcal{T}_{:0+}$ , infinitesimal present  $\Gamma_\delta$ , 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  $b_\mu = \mathbf{I}[\Gamma_\delta; \mathcal{T}_\delta|\mathcal{T}_{:0+}]$  is the rate of change of active information storage. And, the co-information rate  $q_\mu = \mathbf{I}[\mathcal{T}_{:0+}; \Gamma_\delta; \mathcal{T}_\delta]$  is the rate of change of shared information between past, present, and future. These definitions closely parallel those in Ref. [20]

and the *co-information rate* as:

$$q_\mu = \lim_{\delta \rightarrow 0} \frac{d\mathbf{I}[\mathcal{T}_{:0+}; \Gamma_\delta; \mathcal{T}_\delta]}{d\delta} . \tag{12}$$

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

$$\begin{aligned} H_0 &= 2b_\mu + r_\mu + q_\mu , \\ h_\mu &= b_\mu + r_\mu . \end{aligned}$$

So, the entropy rate  $h_\mu$ , the instantaneous rate of information creation, again decomposes into a component  $b_\mu$  that represents active information storage and a component  $r_\mu$  that monitors “dissipated” information. The information-diagram for rates is given in Fig. 7; complementing the causal-state diagram of Fig. 3.

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 noneventually  $\Delta$ -Poisson  $\epsilon$ -machines and, implicitly, their bidirectional machines.

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

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

The normalization factor of this distribution is:

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

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  $\sigma^+$  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, which is defined as  $\mathbf{I}[\mathcal{T}_{0+}; \mathcal{T}_{0-}]$  or equivalently as  $\lim_{T \rightarrow \infty} (H[\mathcal{T}_{0:T}] - h_\mu T)$  [8], where  $\mathcal{T}_{0:T}$  is a future of time length  $T$ , we merely need calculate [10, 11]:

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

Since  $\int_0^\infty \int_0^\infty f(x+y) dx dy = \int_0^\infty xf(x) dx$ , which can be shown by recourse to Riemann sums [16], we have:

$$\mathbf{E} = \int_0^\infty \mu t \phi(t) \log_2(\mu \phi(t)) dt - 2 \int_0^\infty \mu \Phi(t) \log_2(\mu \Phi(t)) dt .$$

This agrees with the formula given in Ref. [13], which was derived by considering the limit of infinitesimal time discretization [22].

Now, we turn to the more technically challenging task of calculating differential information anatomy rates. Suppose that  $\Gamma_\delta$  is a random variable for paths of length  $\delta$ . Each path is uniquely specified by a list of event times. The trajectory distribution is therefore quite complicated. However, only trajectories with zero or one event matter for calculating these differential information anatomy rates. Let  $X_\delta$  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  $h_\mu$ . We can rewrite the path entropy as:

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

For renewal processes, when  $\mu$  is finite, 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:

$$H[\Gamma_\delta|X_\delta] = \Pr(X_\delta = 0) H[\Gamma_\delta|X_\delta = 0] + \Pr(X_\delta = 1) H[\Gamma_\delta|X_\delta = 1] + \Pr(X_\delta = 2) H[\Gamma_\delta|X_\delta = 2].$$

First, we notice that  $\Gamma_\delta$  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  $O(\log \delta)$ . Then, note that  $\Pr(\Gamma_\delta|X_\delta = 2)$  is a probability density function of two variables with the stipulation that neither is negative and that the sum is less than  $\delta$ . Hence, by standard maximum entropy arguments,  $H[\Gamma_\delta|X_\delta = 2]$  is at most  $\log \delta$ . 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:  $H[\Gamma_\delta|X_\delta = 2] \geq H[\Gamma_\delta|X_\delta = 1]$ . The latter, by arguments below, is  $O(\log \delta)$ . Thus, the term  $\Pr(X_\delta = 2) H[\Gamma_\delta|X_\delta = 2]$  is  $O(\delta^2 \log \delta)$  at most.

Finally, to calculate  $H[\Gamma_\delta|X_\delta = 1]$ , 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  $\Pr(\mathcal{T} = t|X_\delta = 1) = \delta^{-1} + O(\delta)$ , and so  $H[\Gamma_\delta|X_\delta = 1]$  is  $\log \delta$  with corrections of  $O(\delta)$ . Hence, the largest corrections to  $H[\Gamma_\delta|X_\delta]$  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:

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

Together, these manipulations give:

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

This then implies:

$$\begin{aligned} H_0 &= \lim_{\delta \rightarrow 0} \frac{dH[\Gamma_\delta]}{d\delta} \\ &= \mu - \mu \log \mu. \end{aligned}$$

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

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

which gives:

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

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

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

This is the nondivergent component of the expression given in Eq. (10) of Ref. [13] for the  $\tau$ -entropy rate of renewal processes, and it agrees with an alternative derivation [23].



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

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

Let us develop a short-time  $\delta$ -asymptotic expansion for  $\Pr(\mathcal{S}_\delta^- = \sigma^- | \mathcal{S}_0^+ = \sigma^+)$ . First, notice we have the Markov chain  $\mathcal{S}_0^+ \rightarrow \mathcal{S}_\delta^+ \rightarrow \mathcal{S}_\delta^-$ , so that:

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

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  $O(\delta^2)$ , so we have:

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

and:

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

Then, straightforward algebra not shown gives:

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

This can be used to derive:

$$b_\mu(\sigma^+) = \frac{\phi(\sigma^+)}{\Phi(\sigma^+)} \left( \log \phi(\sigma^+) - 1 - \int_0^\infty \phi(\sigma^-) \log \phi(\sigma^+ + \sigma^-) d\sigma^- \right),$$

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

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

in nats. From this, we find  $r_\mu$  using:

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

Continuing, we calculate  $q_\mu$  from:

$$\begin{aligned} q_\mu &= H_0 - (h_\mu + b_\mu) \\ &= -\mu \log \mu - \mu + \mu \left( \int_0^\infty \phi(t) \int_0^\infty \phi(t') \log \phi(t + t') dt' dt + 1 \right) \\ &= \mu \int_0^\infty \phi(t) \int_0^\infty \phi(t') \log \phi(t') dt' dt - \mu \log \mu. \end{aligned}$$

And, we calculate  $\rho_\mu$  via:

$$\begin{aligned} \rho_\mu &= H_0 - h_\mu \\ &= -\mu \log \mu - \mu + \mu \int_0^\infty \phi(\sigma^+) \log \phi(\sigma^+) d\sigma^+. \end{aligned}$$

All these quantities are gathered in Table 1, which gives them in bits rather than nats.

## 6 Conclusions

Though the definition of continuous-time causal states of renewal processes parallels that for discrete-time causal states, continuous-time  $\epsilon$ -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  $\epsilon$ -machines do not appear like conventional HMMs—recall Figs. 4, 5, and 6—and most of the information measures (excepting the excess entropy) must be reinterpreted as differential information rates. And so, the machinery required to deploy continuous-time  $\epsilon$ -machines differs significantly from that accompanying discrete-time  $\epsilon$ -machines.

That said, the  $\epsilon$ -machine continuous-time machinery gave us a new way to calculate these information measures. Practically, the formulae in Table 1 provide new approaches to binless plug-in information-measure estimation; e.g., following Ref. [24].

Traditionally, expressions for such information measures come from calculating the time-normalized path entropy of arbitrarily long trajectories; e.g., as in Ref. [25]. Instead, we calculated the path entropy of arbitrarily short trajectories, conditioned on the past. This allows us to extend the results of Ref. [25] for the entropy rate of continuous-time, discrete-alphabet processes to hidden semi-Markov processes; see the sequel Ref. [26]. We hope that our results here pave the way toward understanding the difficulties that lie ahead when studying the structure and information in continuous-time, continuous-value processes.

**Acknowledgements** 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.

## References

1. Shalizi, C.R., Crutchfield, J.P.: Computational mechanics: Pattern and prediction structure and simplicity. *J. Stat. Phys.* **104**, 817–879 (2001)
2. Lohr, W.: Properties of the statistical complexity functional and partially deterministic HMMs. *Entropy* **11**(3), 385–401 (2009)
3. Crutchfield, J.P., Riechers, P., Ellison, C.J.: Exact complexity: Spectral decomposition of intrinsic computation. *Phys. Lett. A* **380**(9–10), 998–1002 (2016)
4. Kelly, D., Dillingham, M., Hudson, A., Wiesner, K.: A new method for inferring hidden Markov models from noisy time sequences. *PLoS ONE* **7**(1), e29703 (2012)
5. Marzen, S., Crutchfield, J.P.: Information anatomy of stochastic equilibria. *Entropy* **16**, 4713–4748 (2014)
6. Riechers, P.M., Crutchfield, J.P.: Beyond the spectral theorem: Decomposing arbitrary functions of non-diagonalizable operators. [arXiv:1607.06526](https://arxiv.org/abs/1607.06526)
7. James, R.G., Burke, K., Crutchfield, J.P.: Chaos forgets and remembers: Measuring information creation, destruction, and storage. *Phys. Lett. A* **378**, 2124–2127 (2014)
8. Crutchfield, J.P., Feldman, D.P.: Regularities unseen, randomness observed: Levels of entropy convergence. *CHAOS* **13**(1), 25–54 (2003)
9. Nair, C., Prabhakar, B., Shah, D.: On entropy for mixtures of discrete and continuous variables. [arXiv:cs/0607075](https://arxiv.org/abs/cs/0607075)
10. Ellison, C.J., Mahoney, J.R., Crutchfield, J.P.: Prediction, retrodiction, and the amount of information stored in the present. *J. Stat. Phys.* **136**(6), 1005–1034 (2009)
11. Crutchfield, J.P., Ellison, C.J., Mahoney, J.R.: Time's barbed arrow: Irreversibility, crypticity, and stored information. *Phys. Rev. Lett.* **103**(9), 094101 (2009)
12. Gaspard, P., Wang, X.-J.: Noise, chaos, and  $(\epsilon, \tau)$ -entropy per unit time. *Phys. Rep* **235**(6), 291–343 (1993)
13. Marzen, S., DeWeese, M.R., Crutchfield, J.P.: Time resolution dependence of information measures for spiking neurons: Scaling and universality. *Front. Comput. Neurosci.* **9**, 109 (2015)
14. Cover, T.M., Thomas, J.A.: *Elements of Information Theory*. Wiley-Interscience, New York (1991)
15. Rao, M., Chen, Y., Vemuri, B.C., Wang, F.: Cumulative residual entropy: a new measure of information. *IEEE Trans. Inform. Theory* **50**(6), 1220–1228 (2004)
16. Marzen, S., Crutchfield, J.P.: Informational and causal architecture of discrete-time renewal processes. *Entropy* **17**(7), 4891–4917 (2015)
17. Lowen, S.B., Teich, M.C.: Fractal renewal processes generate 1/f noise. *Phys. Rev. E* **47**(2), 992–1001 (1993)
18. Marzen, S.E., Crutchfield, J.P.: Statistical signatures of structural organization: The case of long memory in renewal processes. *Phys. Lett. A* **380**(17), 1517–1525 (2016)
19. Risken, H.: *The Fokker–Planck Equation: Methods of Solution and Applications*. Springer Series in Synergetics. Springer, New York (2012)
20. James, R.G., Ellison, C.J., Crutchfield, J.P.: Anatomy of a bit: Information in a time series observation. *CHAOS* **21**(3), 037109 (2011)
21. Yeung, R.W.: A new outlook on Shannon's information measures. *IEEE Trans. Inform. Theory* **37**(3), 466–474 (1991)
22. Pinsker, M.S.: *Information and Information Stability of Random Variables and Processes*. Holden-Day Series in Time Series Analysis. Holden-Day, San Francisco (1964)
23. Girardin, V.: On the different extensions of the ergodic theorem of information theory. In: Baeza-Yates, R., Glaz, J., Gzyl, H., Husler, J., Palacios, J.L. (eds.) *Recent Advances in Applied Probability Theory*, pp. 163–179. Springer, New York (2005)
24. Victor, J.D.: Binless strategies for estimation of information from neural data. *Phys. Rev. E* **66**(5), 051903 (2002)
25. Girardin, V., Linnios, N.: On the entropy for semi-Markov processes. *J. Appl. Prob.* **40**(4), 1060–1068 (2003)
26. Marzen, S., Crutchfield, J.P.: Structure and randomness of continuous-time discrete-event processes. In preparation (2017)