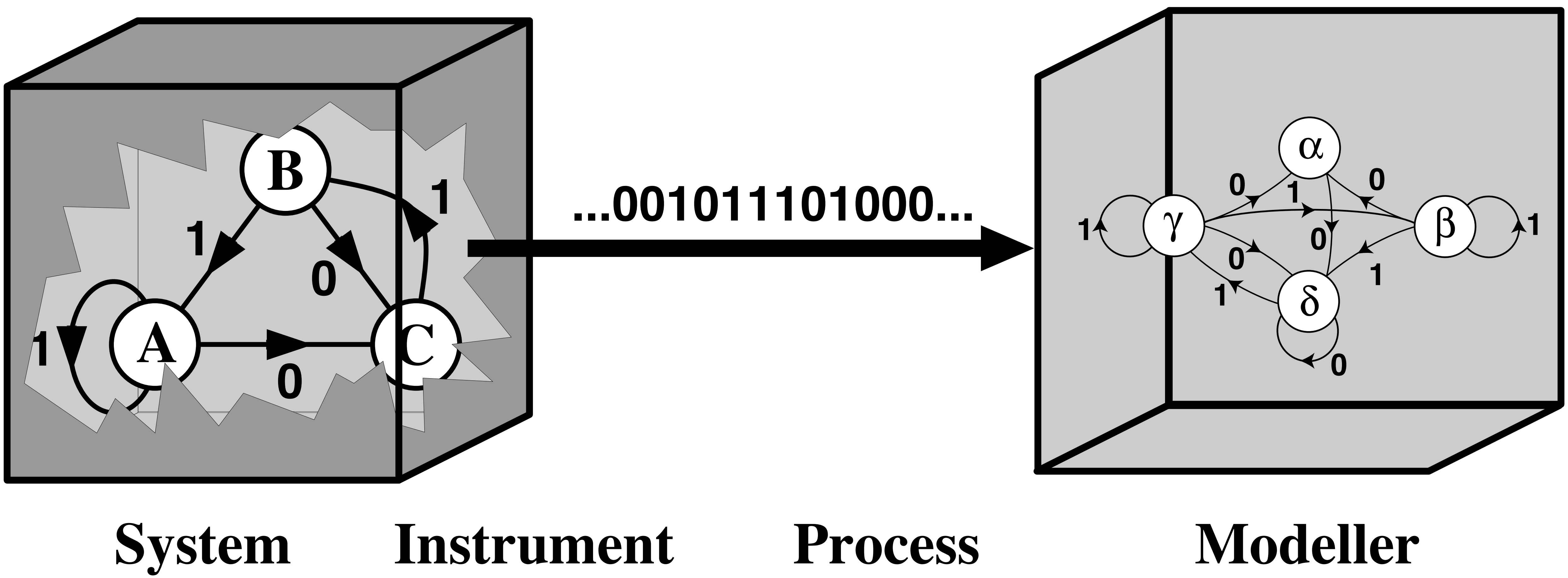
The ϵ-Machine

Reading for this lecture:

CMR article *CMPPSS* and *Lecture Notes*.

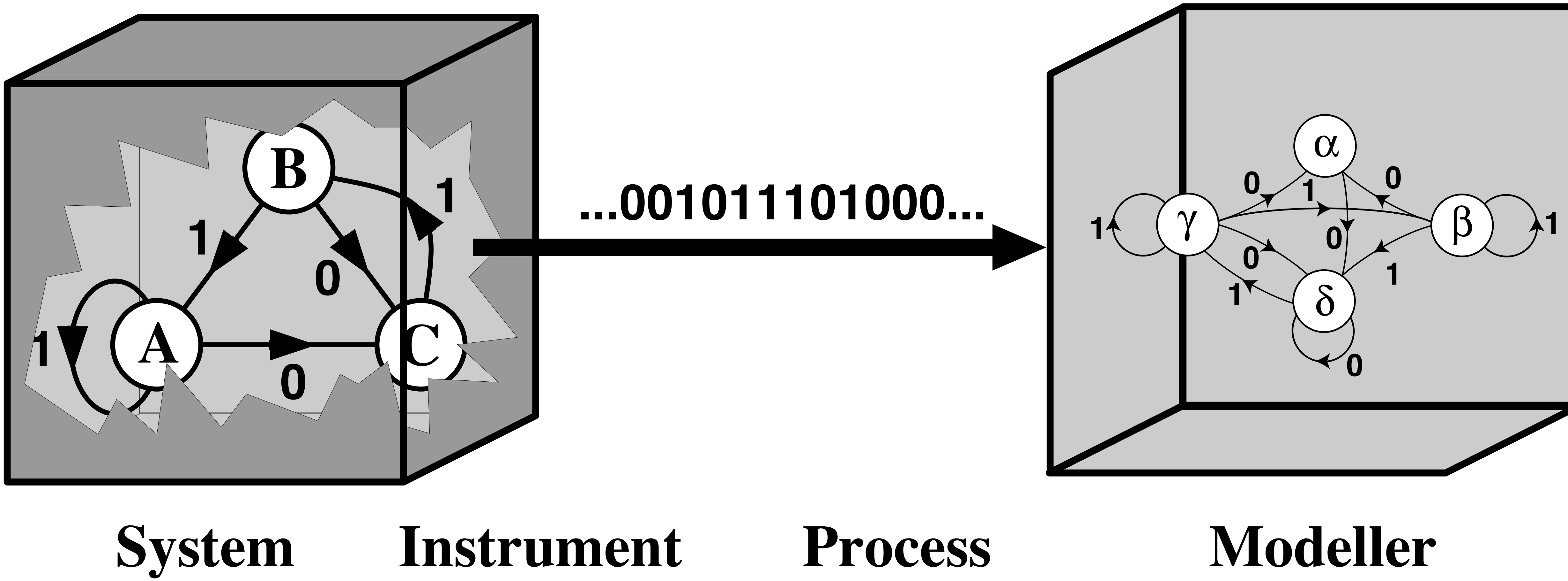
The Learning Channel:



Central questions:

- What are the states?
- What is the dynamic?

The Learning Channel:



Central questions:

What are the states? Causal States

What is the dynamic? The ϵ-Machine

The ϵ-Machine ...

$$\mathcal{M} = \left\{ \boldsymbol{\mathcal{S}}, \{T^{(s)}, s \in \mathcal{A}\} \right\}$$

Causal State:

$$\overleftarrow{s}' \sim \overleftarrow{s}'' \iff \Pr(\overrightarrow{S} \mid \overleftarrow{S} = \overleftarrow{s}') = \Pr(\overrightarrow{S} \mid \overleftarrow{S} = \overleftarrow{s}'')$$

$$\overleftarrow{s}', \overleftarrow{s}'' \in \overleftarrow{\mathbf{S}}$$

Causal state set:

$$\boldsymbol{\mathcal{S}} = \overleftarrow{\mathbf{S}} / \sim \quad = \{\mathcal{S}_0, \mathcal{S}_1, \mathcal{S}_2, \ldots\}$$

Conditional transition probability:

$$T_{ij}^{(s)} = \Pr(\mathcal{S}_j, s | \mathcal{S}_i)$$

$$= \Pr\left(\mathcal{S} = \epsilon(\overleftarrow{s}\, s) | \mathcal{S} = \epsilon(\overleftarrow{s})\right)$$

The ϵ-Machine ...

$$\text{Process} \Rightarrow \text{Predictive equivalence} \Rightarrow \epsilon - \text{Machine}$$

$$\Pr(\overleftrightarrow{S}) \Rightarrow \overleftarrow{\mathbf{S}} / \sim \Rightarrow \mathcal{M} = \left\{ \boldsymbol{\mathcal{S}}, \{T^{(s)}, s \in \mathcal{A}\} \right\}$$

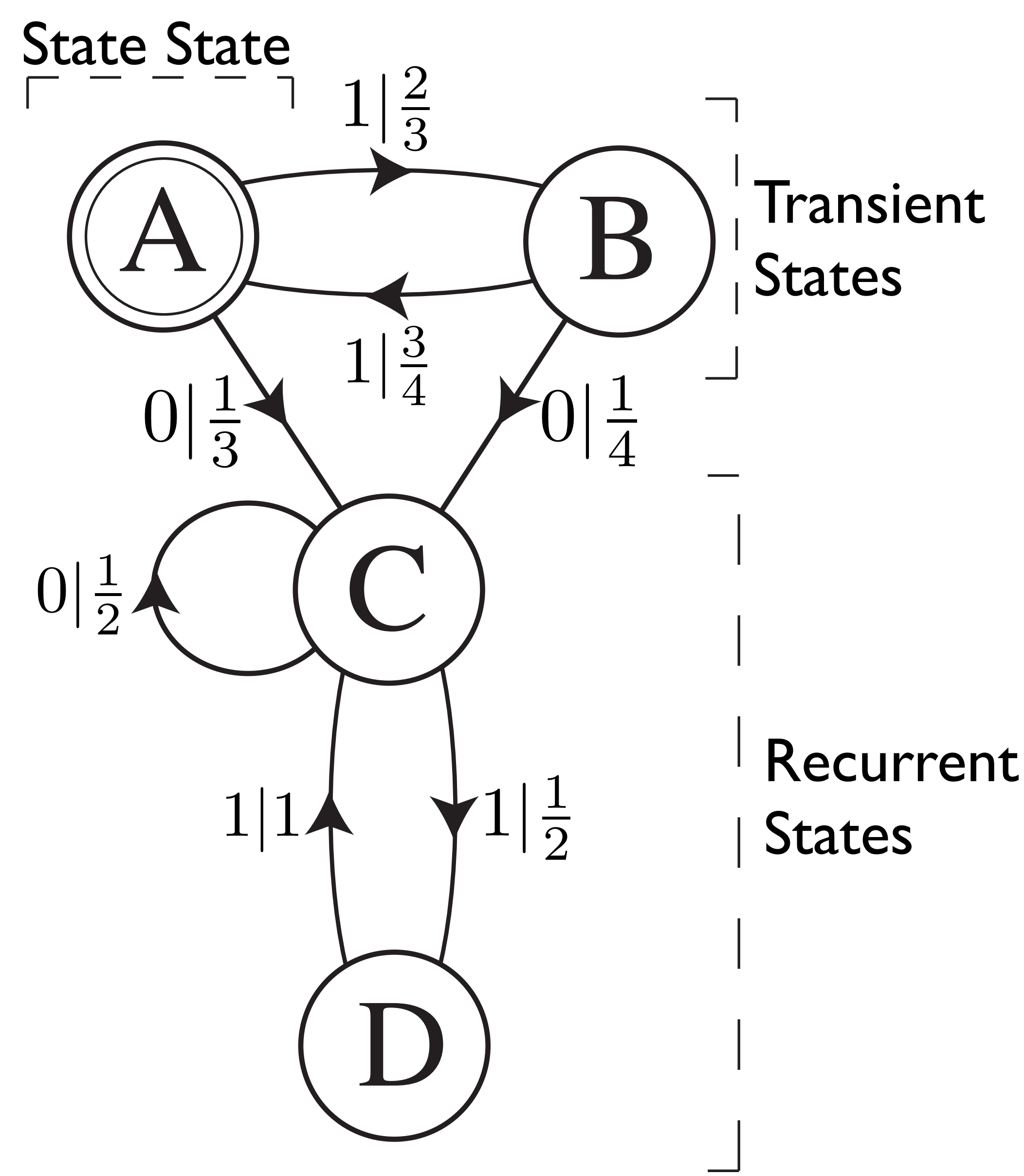
Unique Start State:

$\mathcal{S}_0 = [\lambda]$

$\Pr(\mathcal{S}_0, \mathcal{S}_1, \mathcal{S}_2, \ldots) = (1, 0, 0, \ldots)$

Transient States

Recurrent States



The ϵ-Machine ...

$$\text{Process} \;\Rightarrow\; \text{Predictive equivalence} \;\Rightarrow\; \epsilon - \text{Machine}$$

$$\Pr(\overleftrightarrow{S}) \;\Rightarrow\; \overleftarrow{\mathbf{S}} \,/\sim \;\Rightarrow\; \mathcal{M} = \left\{ \boldsymbol{\mathcal{S}}, \{T^{(s)}, s \in \mathcal{A}\} \right\}$$

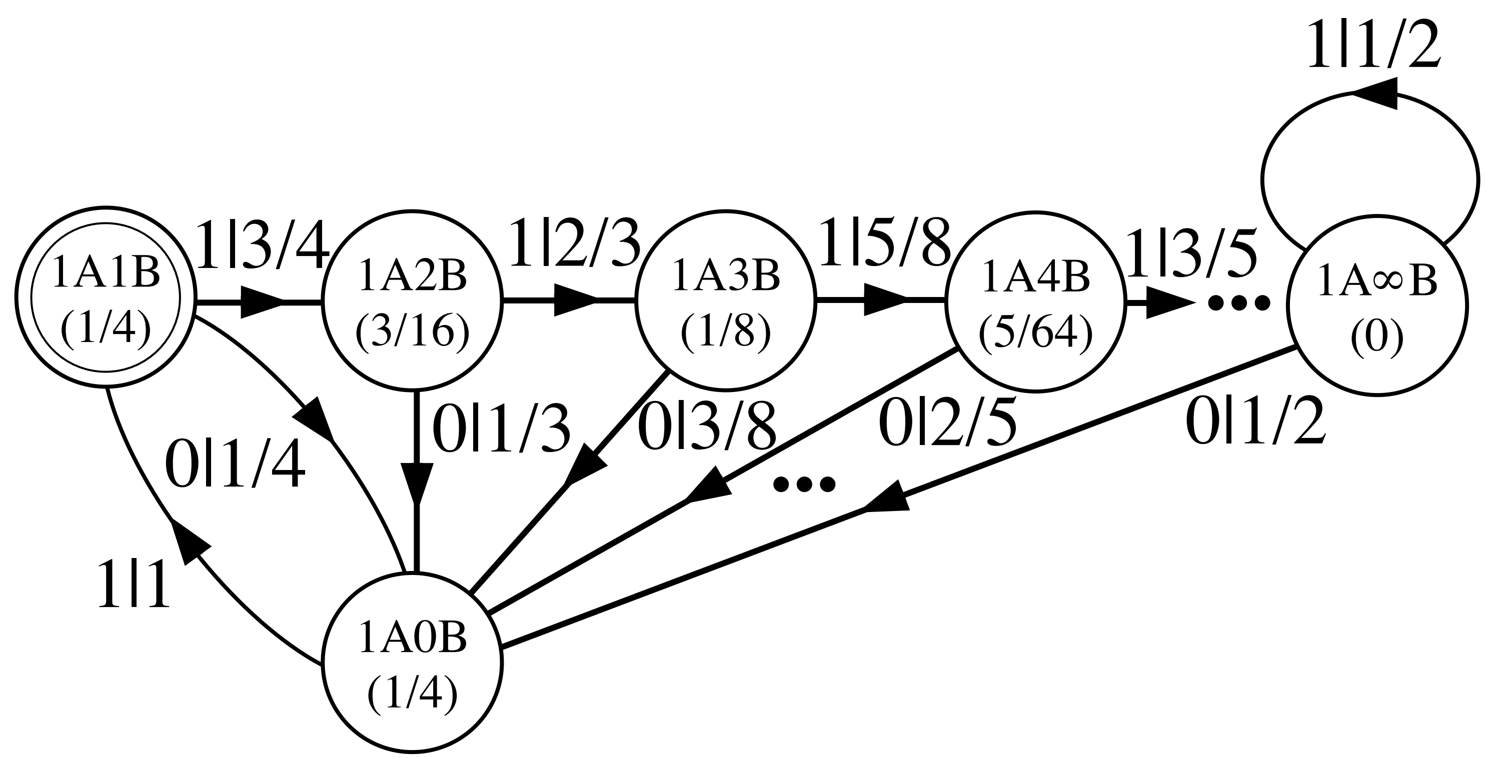
Not always finite state!

The ϵ-Machine ...

$$\text{Process} \;\Rightarrow\; \text{Predictive equivalence} \;\Rightarrow\; \epsilon - \text{Machine}$$

$$\Pr(\overleftrightarrow{S}) \;\Rightarrow\; \overleftarrow{\mathbf{S}} \,/ \sim \;\Rightarrow\; \mathcal{M} = \left\{ \boldsymbol{\mathcal{S}}, \{T^{(s)}, s \in \mathcal{A}\} \right\}$$

Not always finite state!



The ϵ-Machine ...

$$\text{Process} \;\Rightarrow\; \text{Predictive equivalence} \;\Rightarrow\; \epsilon - \text{Machine}$$

$$\Pr(\overleftrightarrow{S}) \;\Rightarrow\; \overleftarrow{\mathbf{S}} \,/\sim \;\Rightarrow\; \mathcal{M} = \left\{ \boldsymbol{\mathcal{S}}, \{T^{(s)}, s \in \mathcal{A}\} \right\}$$

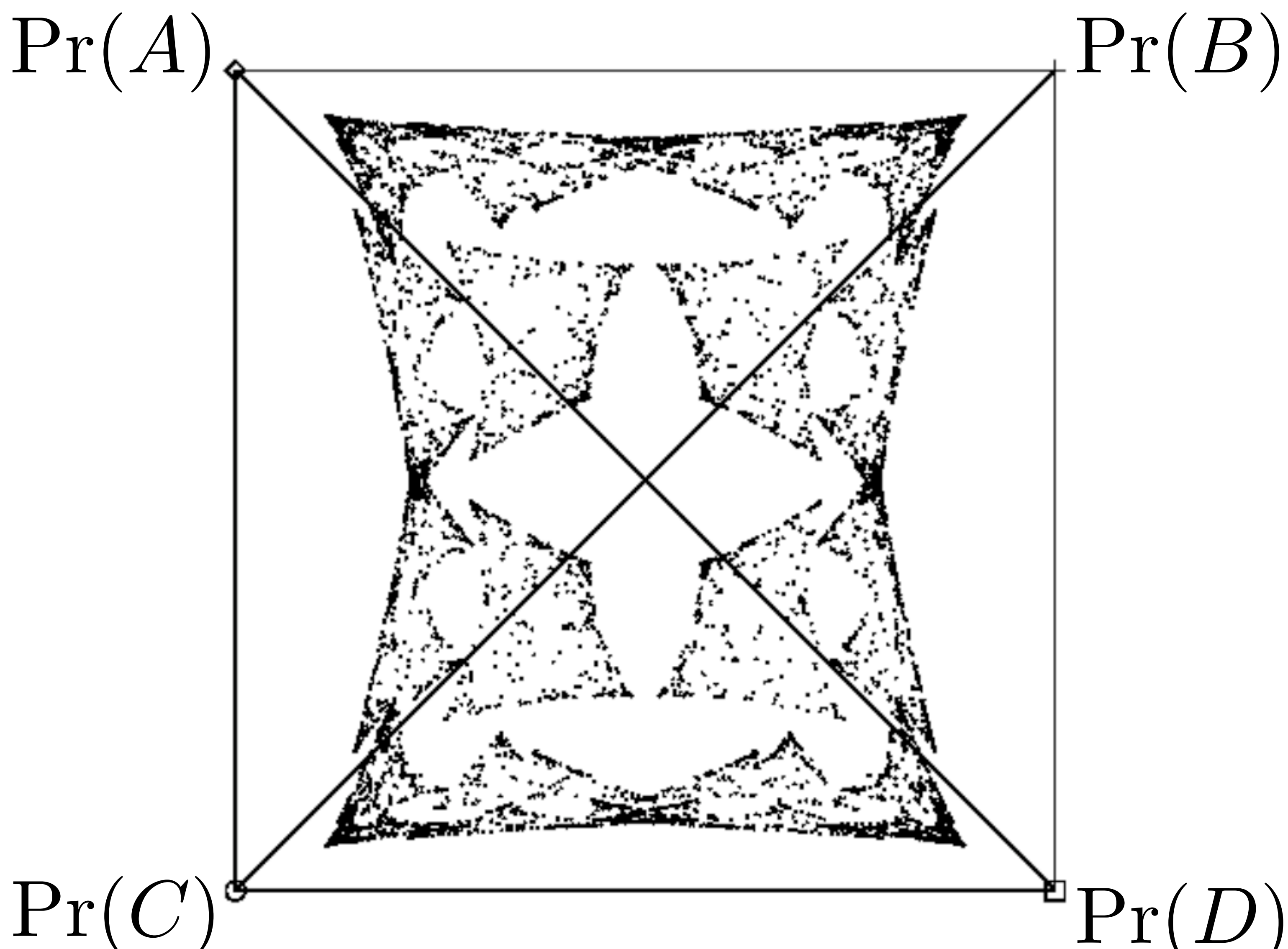
Not always finite state!

The ϵ-Machine ...

$$\text{Process} \;\Rightarrow\; \text{Predictive equivalence} \;\Rightarrow\; \epsilon - \text{Machine}$$

$$\Pr(\overleftrightarrow{S}) \;\Rightarrow\; \overleftarrow{\mathbf{S}} \,/\sim \;\Rightarrow\; \mathcal{M} = \left\{ \boldsymbol{\mathcal{S}}, \{T^{(s)}, s \in \mathcal{A}\} \right\}$$

Not always finite state!



The ϵ-Machine ...

$$\text{Process} \;\Rightarrow\; \text{Predictive equivalence} \;\Rightarrow\; \epsilon - \text{Machine}$$

$$\Pr(\overleftrightarrow{S}) \;\Rightarrow\; \overleftarrow{\mathbf{S}} \,/ \sim \;\Rightarrow\; \mathcal{M} = \left\{ \boldsymbol{\mathcal{S}}, \{T^{(s)}, s \in \mathcal{A}\} \right\}$$

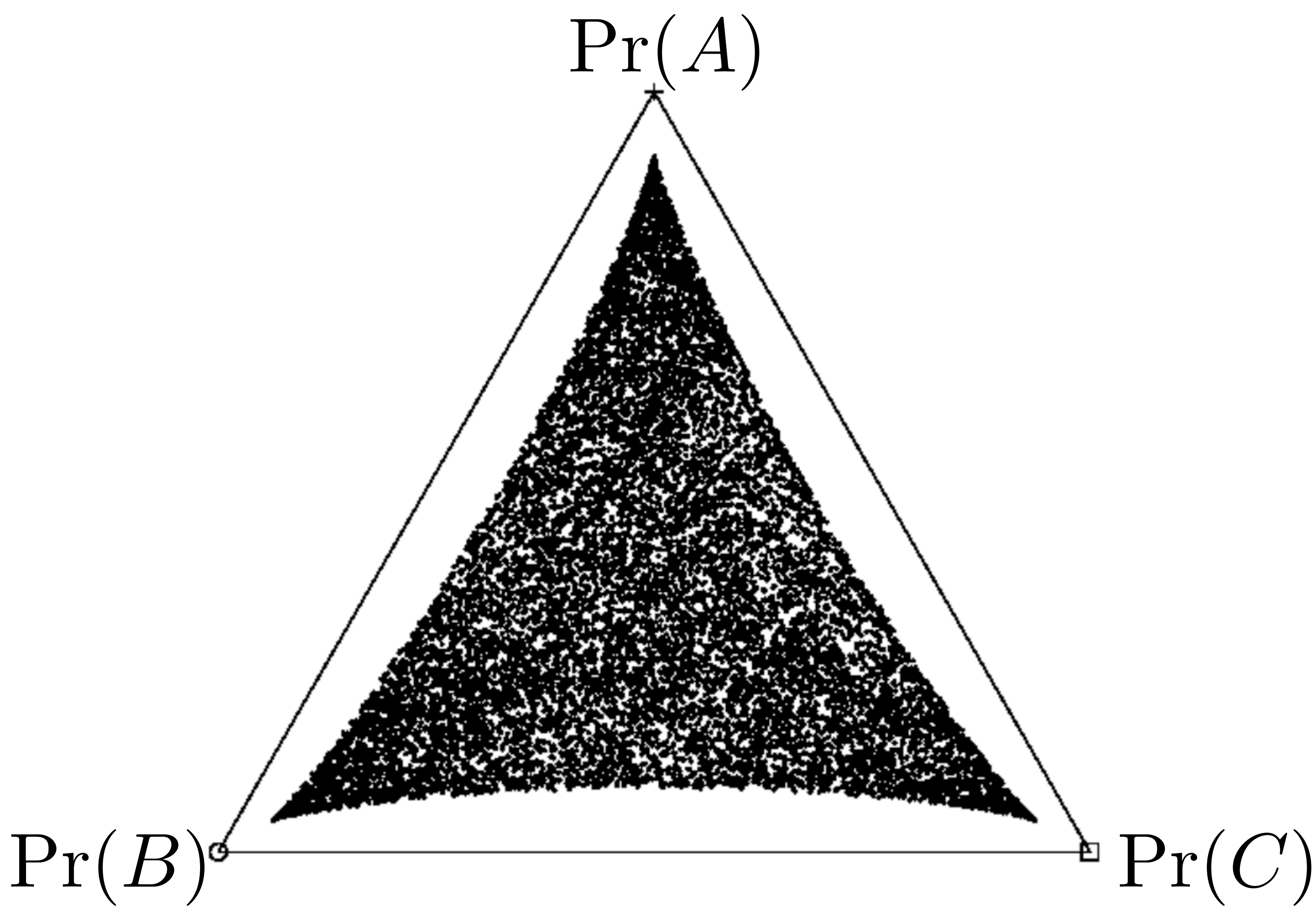
Not always finite state!

The ϵ-Machine ...

$$\text{Process} \;\Rightarrow\; \text{Predictive equivalence} \;\Rightarrow\; \epsilon - \text{Machine}$$

$$\Pr(\overleftrightarrow{S}) \;\Rightarrow\; \overleftarrow{\mathbf{S}} \,/\sim \;\Rightarrow\; \mathcal{M} = \left\{ \boldsymbol{\mathcal{S}}, \{T^{(s)}, s \in \mathcal{A}\} \right\}$$

Not always finite state!



The ϵ-Machine ...

A Model of a Process $\Pr(\overleftrightarrow{S})$:

ϵ-Machine reproduces the process's word distribution:

$$\Pr(s^1), \Pr(s^2), \Pr(s^3), \ldots$$

$$s^L = s_1 s_2 \ldots s_L \qquad \mathcal{S}(t=0) = \mathcal{S}_0$$

$$\begin{aligned} \Pr(s^L) \quad &= \quad \Pr(\mathcal{S}_0)\Pr(\mathcal{S}_0 \rightarrow_{s=s_1} \mathcal{S}(1))\Pr(\mathcal{S}(1) \rightarrow_{s=s_2} \mathcal{S}(2)) \\ & \quad\quad \cdots \Pr(\mathcal{S}(L-1) \rightarrow_{s=s_L} \mathcal{S}(L)) \end{aligned}$$

Initially, $\Pr(\mathcal{S}_0) = 1.$

$$\Pr(s^L) = \prod_{l=1}^{L} T^{(s_l)}_{i=\epsilon(s^{l-1}),j=\epsilon(s^l)}$$

The ϵ-Machine ...

A Model of a Process $\Pr(\overleftrightarrow{S})$ ...

Calculate word distribution from *recurrent states:* $\mathcal{S}_i \in \mathcal{S}_{\text{recurrent}}$

$$\Pr(s^1), \Pr(s^2), \Pr(s^3), \ldots$$

Get $\langle \pi | = (p_{\mathcal{S}_1}, p_{\mathcal{S}_2}, \ldots)$ from $T = \sum_{s \in \mathcal{A}} T^{(s)}$

Then

$$\Pr(s) = \langle \pi | T^{(s)} | 1 \rangle$$

$$| 1 \rangle = \begin{pmatrix} 1 \\ 1 \\ \vdots \end{pmatrix}$$

$$\Pr(s_0 s_1) = \langle \pi | T^{(s_0)} T^{(s_1)} | 1 \rangle$$

...

$$\Pr(s^L) = \langle \pi | T^{(s^L)} | 1 \rangle$$

$$T^{(s^L)} = T^{(s_0)} T^{(s_1)} \ldots T^{(s_{L-1})}$$

The ϵ-Machine ...

Properties:

- Shielding: Conditional independence of future & past
- Unifilar
- Markovian
- Optimal predictor
- Minimal size
- Unique

The ϵ-Machine ...

Causal shielding:

Past and future are independent given causal state: $\overleftarrow{S} \perp_{\mathcal{S}} \overrightarrow{S}$

Process: $\Pr(\overleftrightarrow{S}) = \Pr(\overleftarrow{S}\,\overrightarrow{S})$

$$\Pr(\overleftarrow{S}\,\overrightarrow{S}\,|\mathcal{S}) = \Pr(\overleftarrow{S}\,|\mathcal{S})\,\Pr(\overrightarrow{S}\,|\mathcal{S})$$

Causal states shield past & future from each other.

Similar to states of a Markov chain, but for hidden processes.

In fact, there is a Markov chain (in info-theoretic sense):

$$\overleftarrow{X} \Rightarrow \mathcal{S} \Rightarrow \overrightarrow{X}$$

The ϵ-Machine ...

Proof sketch:

$$\begin{aligned} \Pr(\overleftrightarrow{S} | \mathcal{S}) &= \Pr(\overleftarrow{S}\overrightarrow{S} | \mathcal{S}) \\ &= \Pr(\overrightarrow{S} \mid \overleftarrow{S}, \mathcal{S}) \Pr(\overleftarrow{S} | \mathcal{S}) \end{aligned}$$

Will show: $\Pr(\overrightarrow{S} \mid \overleftarrow{S}, \mathcal{S}) = \Pr(\overrightarrow{S} | \mathcal{S})$

$\mathcal{S} = \epsilon(\overleftarrow{s}) \Rightarrow$

$$\Pr\left(\overrightarrow{S} | \overleftarrow{S} = \overleftarrow{s}', \mathcal{S} = \epsilon(\overleftarrow{s})\right) = \Pr\left(\overrightarrow{S} | \overleftarrow{S} = \overleftarrow{s}\right) \quad (\overleftarrow{s}' \in [\overleftarrow{s}])$$

But, also, $\Pr\left(\overrightarrow{S} \mid \overleftarrow{S} = \overleftarrow{s}\right) = \Pr\left(\overrightarrow{S} | \mathcal{S} = \epsilon(\overleftarrow{s})\right)$ (Causal equiv. rel'n)

So, $\Pr\left(\overrightarrow{S} \mid \overleftarrow{S} = \overleftarrow{s}, \mathcal{S} = \sigma\right) = \Pr\left(\overrightarrow{S} | \mathcal{S} = \sigma\right)$ $\boxed{\circlearrowleft}$

The ϵ-Machine ...

$\epsilon\mathrm{Ms}$ are Unifilar: $(\mathcal{S}_t, s) \rightarrow$ unique $\mathcal{S}_{t+1}$

(in automata theory, "deterministic")

That is:

(1) $\mathcal{S}_i \in \boldsymbol{\mathcal{S}}, \; s \in \mathcal{A}$, at most one $\mathcal{S}_j \in \boldsymbol{\mathcal{S}}$:

$$\overleftarrow{s} \in \mathcal{S}_i \Rightarrow \overleftarrow{s}\, s \in \mathcal{S}_j$$

(2) If there is a next causal state j:

$$\mathcal{S}_{k \neq j} \in \boldsymbol{\mathcal{S}} \Rightarrow T_{ik}^{(s)} = 0$$

(3) If there is not:

$$T_{ij}^{(s)} = 0$$

The ϵ-Machine ...

Unifilarity ...

Proof sketch:

Must show $\overleftarrow{s} \sim \overleftarrow{s}' \Rightarrow \overleftarrow{s}\, s \sim \overleftarrow{s}'\, s$

Futures with symbol prefixed: sF $\qquad F \subseteq \mathcal{A}^{\mathbb{Z}^+}$

$$\overleftarrow{s} \sim \overleftarrow{s}' \quad \Rightarrow \quad \Pr\left(\overrightarrow{S} \in sF | \overleftarrow{S} = \overleftarrow{s}\right) = \Pr\left(\overrightarrow{S} \in sF | \overleftarrow{S} = \overleftarrow{s}'\right)$$

$$\Pr\left(\overrightarrow{S}^1 = s, \overrightarrow{S}_1 \in F | \overleftarrow{S} = \overleftarrow{s}\right) = \Pr\left(\overrightarrow{S}^1 = s, \overrightarrow{S}_1 \in F | \overleftarrow{S} = \overleftarrow{s}'\right)$$

$\Pr(X, Y|Z) = \Pr(Y|Z)\Pr(X|Y, Z)$

$$\Pr\left(\overrightarrow{S}_1 \in F | \overrightarrow{S}^1 = s, \overleftarrow{S} = \overleftarrow{s}\right) \Pr\left(\overrightarrow{S}^1 = s | \overleftarrow{S} = \overleftarrow{s}\right) = \Pr\left(\overrightarrow{S}_1 \in F | \overrightarrow{S}^1 = s, \overleftarrow{S} = \overleftarrow{s}'\right) \Pr\left(\overrightarrow{S}^1 = s | \overleftarrow{S} = \overleftarrow{s}'\right)$$

$$\Pr\left(\overrightarrow{S}_1 \in F | \overleftarrow{S} = \overleftarrow{s}\, s\right) = \Pr\left(\overrightarrow{S}_1 \in F | \overleftarrow{S} = \overleftarrow{s}'\, s\right)$$

(Stationarity and
by assumption
$\Pr(\overrightarrow{S}^1 = s | \overleftarrow{s}) = 1$
$\Pr(\overrightarrow{S}^1 = s | \overleftarrow{s}') = 1$)

$$\Rightarrow \overleftarrow{s}\, s \sim \overleftarrow{s}'\, s \qquad \boxed{\circlearrowleft}$$

The ϵ-Machine ...

Unifiliarity ...

Consequence:

Unifilarity: 1-1 map between state-sequences & symbol-sequences.

Entropy rate expression requires this 1-1 mapping.

Can (must) use $\epsilon \mathrm{M}$ to calculate entropy rate h_μ.

The ϵ-Machine ...

ϵMs are first-order Markov in state sequences:

$$\Pr(\mathcal{S}_t | \ldots \mathcal{S}_{t-2} \mathcal{S}_{t-1}) = \Pr(\mathcal{S}_t | \mathcal{S}_{t-1})$$

Proof sketch:

Show

$$\Pr(\mathcal{S}_t | \mathcal{S}_{t-2} \mathcal{S}_{t-1}) = \Pr(\mathcal{S}_t | \mathcal{S}_{t-1})$$

(Additional conditioning removed by induction.)

$$\Pr(\mathcal{S}_t \in M \subset \boldsymbol{\mathcal{S}} | \mathcal{S}_{t-2} \mathcal{S}_{t-1}) = \Pr(\overrightarrow{S}^1 \in A \subset \mathcal{A} | \mathcal{S}_{t-2} \mathcal{S}_{t-1})$$

$$= \Pr(\overrightarrow{S}^1 \in A | \mathcal{S}_{t-1}) \quad \text{(Causal shielding)}$$

$$= \Pr(\mathcal{S}_t \in M | \mathcal{S}_{t-1}) \quad \boxed{\circlearrowleft}$$

The ϵ-Machine ...

ϵMs are Optimal Predictors:

Compared to any rival effective states R :

$$H\left[\overrightarrow{S}^L | R\right] \geq H\left[\overrightarrow{S}^L | \mathcal{S}\right]$$

Proof sketch: $H\left[\overrightarrow{S}^L | \mathcal{S}\right] = H\left[\overrightarrow{S}^L | \overleftarrow{s} \in \mathcal{S}\right]$ (Causal equiv. rel'n)

$$= H\left[\overrightarrow{S}^L | \overleftarrow{s}\right]$$

$$\leq H\left[\overrightarrow{S}^L | R\right]$$

$$R = \eta(\overleftarrow{s})$$

(Data processing inequality)

$\square$

The ϵ-Machine ...

ϵMs are Optimal Predictors ...

Lemma:

$$h_\mu(\mathcal{S}) = h_\mu$$

Proof:

$$\begin{aligned} h_\mu(\mathcal{S}) &= \lim_{L\to\infty} \frac{1}{L} H\left[\overrightarrow{S}^L | \mathcal{S}\right] && \text{(Block entropy)} \\ &= \lim_{L\to\infty} \frac{1}{L} H\left[\overrightarrow{S}^L | \overleftarrow{S}\right] && \text{(Causal equiv. rel'n)} \\ &= \lim_{L\to\infty} \frac{1}{L} L H\left[S | \overleftarrow{S}\right] && \text{(Stationarity)} \\ &= H\left[S | \overleftarrow{S}\right] \\ &= h_\mu \end{aligned}$$

$\circlearrowleft$ ∎

Corollary: $h_\mu(R) \geq h_\mu$

The ϵ-Machine ...

ϵMs are Optimal Predictors ...

Corollary (Maximal Prescience): $\Pi(R) \leq \Pi(\mathcal{S})$

Rival model: $\Pi(R) = \log_2 |\mathcal{A}| - h_\mu(R)$

But: $\Pi(\mathcal{S}) = \log_2 |\mathcal{A}| - h_\mu = \mathbf{G}$

So: $\Pi(R) \leq \Pi(\mathcal{S})$

$h_\mu(R) \geq h_\mu$

The ϵ-Machine ...

ϵMs are Optimal Predictors ...

Remarks:

(1) Causal states contain every difference (in past)
that makes a difference (to future)
(Recall Bateson "information")

(2) Causal states are sufficient statistics for the future.
(See below.)

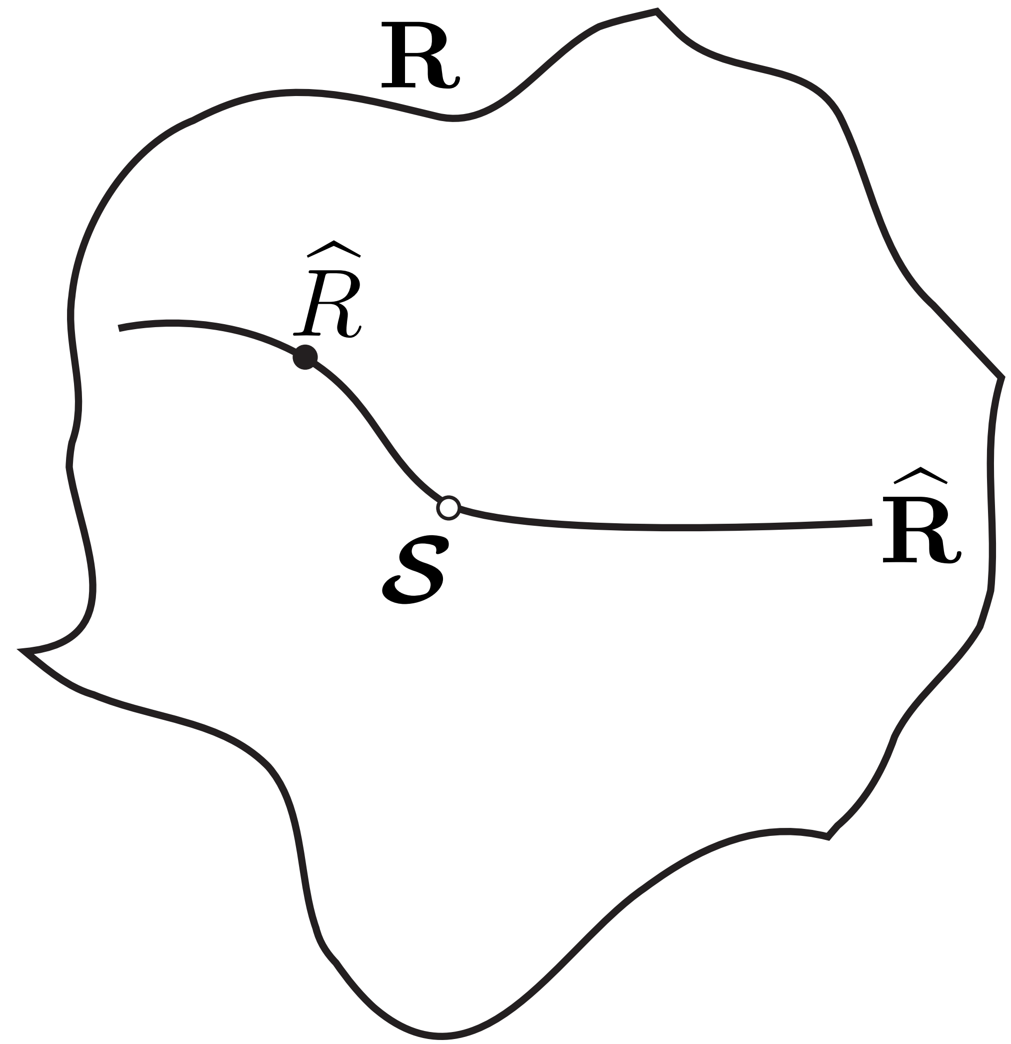
The ϵ-Machine ...

Prescient Rivals $\widehat{\mathbf{R}}$:

Alternative models that are optimal predictors

$\widehat{R} \in \widehat{\mathbf{R}}$

$$H[\overrightarrow{S}^L | \widehat{R}] = H[\overrightarrow{S}^L | \mathcal{S}]$$



(Prescient rivals are sufficient statistics for process's future.)

The ϵ-Machine ...

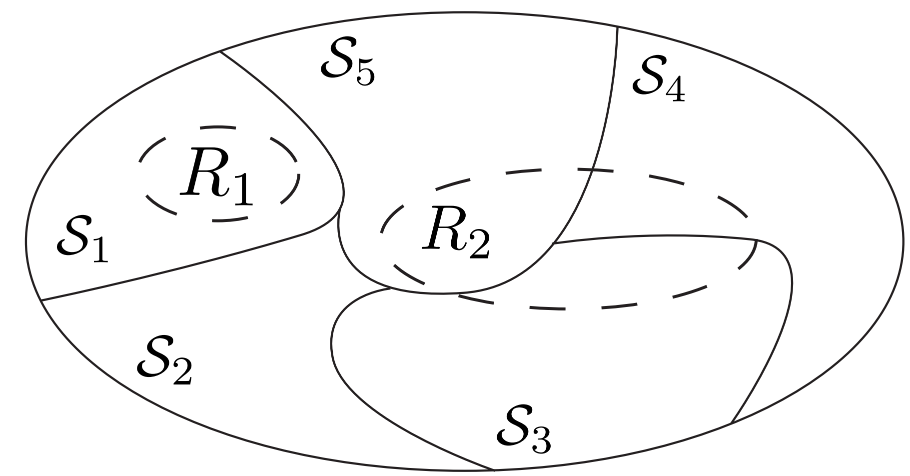
Prescient rivals are refinements of causal states:

Proof sketch:

(1) Either $R_k \subseteq \mathcal{S}_i$:

Then make same prediction,

$$\Pr\left(\overrightarrow{S}|R_k\right) = \Pr\left(\overrightarrow{S}|\mathcal{S}_i\right)$$



(2) Or R_k consists of pieces of various $\mathcal{S}_i$.

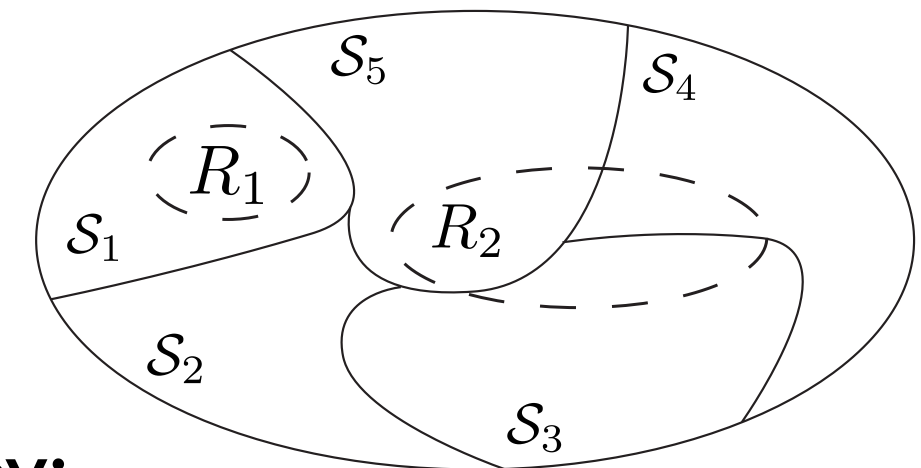
(Not a refinement.)

Then its morph is a statistical mixture of various $\mathcal{S}_i$ morphs:

$$\Pr(\overrightarrow{S}|R_k) = \sum_i c_i \Pr(\overrightarrow{S}|\mathcal{S}_i)$$

The ϵ-Machine ...

Prescient rivals are refinements of causal states:



Proof sketch ...

But mixing distributions increases entropy:

$$H\left[\sum_i c_i P_i\right] \geq \sum_i H(P_i)$$

Thus, worse prediction with rival.

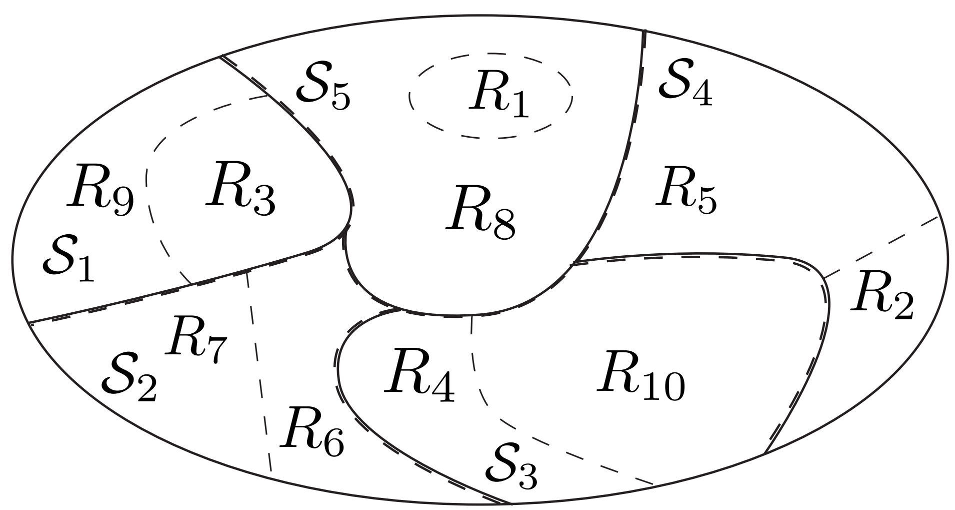
Contradiction!

The ϵ-Machine ...

Prescient rivals are refinements of causal states ...

Proof sketch ...

To be equally prescient, rival must be a refinement:



The ϵ-Machine ...

Minimal Statistical Complexity:

For all prescient rivals, $\epsilon\mathrm{M}$ is the smallest:

$$C_\mu(\widehat{R}) \geq C_\mu(\mathcal{S})$$

Proof sketch:

(1) Prescient rivals are refinements, so

$$\exists g : \mathcal{S} = g(\widehat{R})$$

(2) But

$$H[f(X)] \leq H[X] \Rightarrow H[\mathcal{S}] = H[g(\widehat{R})] \leq H[\widehat{R}]$$

(3) So $C_\mu \leq H[\widehat{R}]$

The ϵ-Machine ...

Minimal Statistical Complexity ...

Consequence:

(1) C_μ measures historical information process stores.

(2) This would not be true, if not minimal representation.

The ϵ-Machine ...

$\epsilon\mathrm{Ms}$ are Unique:

Prescient rival of same size is, up to state relabeling, the $\epsilon\mathrm{M}$.

$$C_\mu(\widehat{R}) = C_\mu(\mathcal{S}) \Rightarrow \widehat{R} = \mathcal{S}$$

Proof Sketch:

(1) Refinement: $\mathcal{S} = g(\widehat{R})$

(2) Other way? $f : \widehat{R} = f(\mathcal{S})$

(3) Show $H[\widehat{R}|\mathcal{S}] = 0$. Consider: $I[\mathcal{S};\widehat{\mathcal{R}}]$

$$H[\mathcal{S}] - H[\mathcal{S}|\widehat{R}] = H[\widehat{R}] - H[\widehat{R}|\mathcal{S}]$$

$$H[\mathcal{S}] = H[\widehat{R}] - H[\widehat{R}|\mathcal{S}] \qquad H[\mathcal{S}|\widehat{R}] = 0$$

But $H[\mathcal{S}] = H[\widehat{R}]$

So $H[\widehat{R}|\mathcal{S}] = 0$

$$\Rightarrow \widehat{R} = f(\mathcal{S})$$

$$\Rightarrow g = f^{-1}$$

The ϵ-Machine ...

$\epsilon\mathrm{Ms}$ have Minimal State-Stochasticity:

$$H[\widehat{\mathcal{R}}_t|\widehat{\mathcal{R}}_{t-1}] \geq H[\mathcal{S}_t|\mathcal{S}_{t-1}]$$

Proof Sketch:

(1) Entropy Chain Rule: $H[X, Y|Z] = H[Y|Z] + H[X|Y, Z]$

$$H\left[\mathcal{S}_t, \overrightarrow{S}^1|\mathcal{S}_{t-1}\right] = H\left[\overrightarrow{S}^1|\mathcal{S}_{t-1}\right] + H\left[\mathcal{S}_t|, \overrightarrow{S}^1, \mathcal{S}_{t-1}\right]$$

(2) Unfilarity:

$$H\left[\mathcal{S}_t|\mathcal{S}_{t-1}, \overrightarrow{S}^1\right] = 0$$

(3) So:

$$H\left[\overrightarrow{S}^1 |\mathcal{S}_{t-1}\right] = H\left[\mathcal{S}_t, \overrightarrow{S}^1 |\mathcal{S}_{t-1}\right]$$

The ϵ-Machine ...

ϵMs have Minimal State-Stochasticity ...

(4) Again:

$$H\left[\widehat{\mathcal{R}}_t, \overrightarrow{S}^1 | \widehat{\mathcal{R}}_{t-1}\right] = H\left[\overrightarrow{S}^1 | \widehat{\mathcal{R}}_{t-1}\right] + H\left[\widehat{\mathcal{R}}_t | \overrightarrow{S}^1, \widehat{\mathcal{R}}_{t-1}\right]$$

$$\geq H\left[\overrightarrow{S}^1 | \widehat{\mathcal{R}}_{t-1}\right]$$

$$= H\left[\overrightarrow{S}^1 | \mathcal{S}_{t-1}\right] \quad \text{(Refinement)}$$

$$= H\left[\mathcal{S}_t, \overrightarrow{S}^1 | \mathcal{S}_{t-1}\right]$$

(5) Also, by chain rule:

$$H\left[\widehat{\mathcal{R}}_t, \overrightarrow{S}^1 | \widehat{\mathcal{R}}_{t-1}\right] = H\left[\widehat{\mathcal{R}}_t | \widehat{\mathcal{R}}_{t-1}\right] + H\left[\overrightarrow{S}^1 | \widehat{\mathcal{R}}_t, \widehat{\mathcal{R}}_{t-1}\right]$$

(6) Putting (4) and (5) together gives:

$$H\left[\widehat{\mathcal{R}}_t | \widehat{\mathcal{R}}_{t-1}\right] + H\left[\overrightarrow{S}^1 | \widehat{\mathcal{R}}_t, \widehat{\mathcal{R}}_{t-1}\right] \geq H\left[\mathcal{S}_t, \overrightarrow{S}^1 | \mathcal{S}_{t-1}\right]$$

The ϵ-Machine ...

ϵMs have Minimal State-Stochasticity ...

(7) Expand RHS of (6) and re-arrange:

$$H[\widehat{\mathcal{R}}_t|\widehat{\mathcal{R}}_{t-1}] - H[\mathcal{S}_t|\mathcal{S}_{t-1}] \geq H\left[\overrightarrow{S}^1 |\mathcal{S}_t, \mathcal{S}_{t-1}\right] - H\left[\overrightarrow{S}^1 |\widehat{\mathcal{R}}_t, \widehat{\mathcal{R}}_{t-1}\right]$$

(8) Note:

$$\mathcal{S}_t = g(\widehat{\mathcal{R}}_t) \Rightarrow (\mathcal{S}_t, \mathcal{S}_{t-1}) = g'(\widehat{\mathcal{R}}_t, \widehat{\mathcal{R}}_{t-1})$$

(9) So,

$$H\left[\overrightarrow{S}^1 |\widehat{\mathcal{R}}_t, \widehat{\mathcal{R}}_{t-1}\right] \leq H\left[\overrightarrow{S}^1 |\mathcal{S}_t, \mathcal{S}_{t-1}\right] \qquad H[X|Y] \leq H[X|g(Y)]$$

(10) And so RHS of (7) > 0 and we have:

$$H[\widehat{\mathcal{R}}_t|\widehat{\mathcal{R}}_{t-1}] \geq H[\mathcal{S}_t|\mathcal{S}_{t-1}]$$

The ϵ-Machine ...

Random variable $X \sim \Pr_\theta(x)$

Sufficient statistic $T(X)$ for θ : [*EIT*, Section 2.9]

Contains all info in X for θ.

That is, $I[\theta; X] = I[\theta; T(X)]$

Minimal sufficient statistic:

$T(X)$ is a function of every other sufficient statistic $U(X)$.

The ϵ-Machine ...

$\epsilon\mathrm{M}$ is a Minimal Sufficient Statistic for a Process.

Proof Sketch:

(1) Maximal prescience gives sufficiency:

$$I[\overrightarrow{S}^L; \mathcal{S}] = I[\overrightarrow{S}^L; \overleftarrow{S}]$$

(2) In fact, every prescient rival $\widehat{\mathcal{R}}$ is a sufficient statistic.

$$I[\overrightarrow{S}^L; \widehat{\mathcal{R}}] = I[\overrightarrow{S}^L; \overleftarrow{S}]$$

(3) $\epsilon\mathrm{M}$ is minimal sufficient statistic:
Rival states are refinements of causal states: $\mathcal{S} = g(\widehat{\mathcal{R}})$. $\boxed{\circlearrowleft}$

Lesson: You can calculate everything about process from its $\epsilon\mathrm{M}$.

The ϵ-Machine ...

Summary:

$\epsilon\mathrm{M}$:

(1) Optimal predictor: Lower prediction error than any rival.

(2) Minimal size: Smallest of the prescient rivals.

(3) Unique: Any smallest, optimal, unifilar predictor is equivalent.

(4) Model of the process: Reproduces all of process's statistics.

(5) Causal shielding: Renders process's future independent of past.

The ϵ-Machine ...

Dynamical system's intrinsic computation:

(1) How much of past does process store?

(2) In what architecture is that information stored?

(3) How is stored information used to produce future behavior?

The ϵ-Machine ...

Reading for next lecture:

CMR articles *CMPPSS & RURO*.