Measures of Structural Complexity

Reading for this lecture:

CMR articles *CMPPSS* and *RURO* and *Lecture Notes*.

Measures of Complexity ...

Information Measures of Complexity:

Measures		Interpretation
Entropy Rate	h_μ	Intrinsic Randomness
Excess Entropy	$\mathbf{E}$	Info: Past to Future
Predictability Gain	$\mathbf{G}$	Redundancy
Transient Information	$\mathbf{T}$	Synchronization

How related to stored information (statistical complexity)?

How to get from $\epsilon\mathrm{M}$?

Measures of Complexity ...

Measures from the $\epsilon\mathrm{M}$:

Entropy Rate of a Process:

$$h_\mu(\Pr(\overleftrightarrow{S})) = \lim_{L \to \infty} \frac{H(L)}{L}$$

Directly from process's $\epsilon\mathrm{M}$:

$$h_\mu\left(\Pr(\overleftrightarrow{S})\right) = h_\mu(\boldsymbol{\mathcal{S}})$$

Measures of Complexity ...

Measures from the $\epsilon\mathrm{M}$:

Entropy Rate given $\epsilon\mathrm{M}$:

$$h_\mu(\boldsymbol{\mathcal{S}}) = -\sum_{\mathcal{S}\in\boldsymbol{\mathcal{S}}} \Pr(\mathcal{S}) \sum_{s\in\mathcal{A},\mathcal{S}'\in\boldsymbol{\mathcal{S}}} T^{(s)}_{\mathcal{S}\mathcal{S}'} \log_2 T^{(s)}_{\mathcal{S}\mathcal{S}'}$$

where $\Pr(\mathcal{S})$ is casual-state asymptotic probability.

Possible only due to $\epsilon\mathrm{M}$ unifilarity!

1-1 mapping between measurement sequences & internal paths.

Measures of Complexity ...

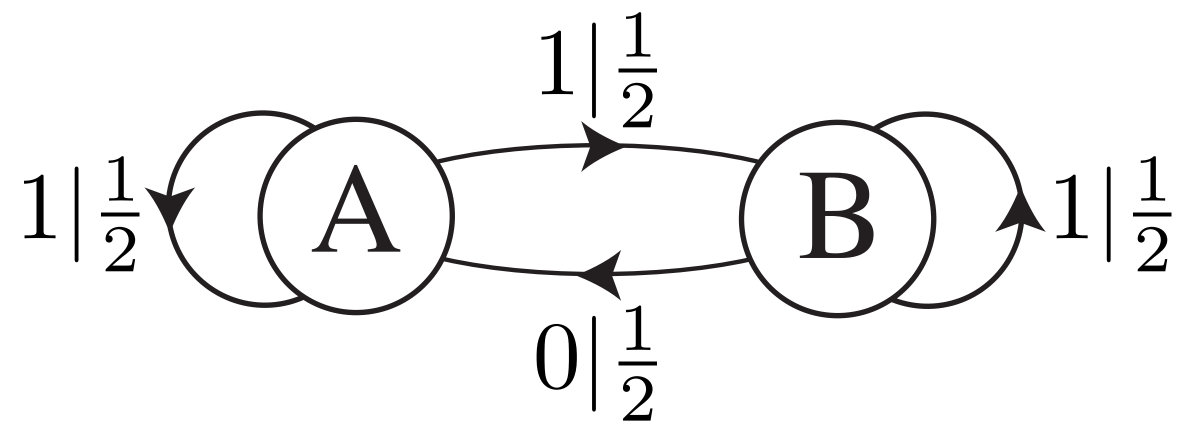
Measures from the $\epsilon\mathrm{M}$:

Entropy rate ...

Possible only due to unifilarity ... Why?

What if you use a nonunifilar HMM presentation?

Consider nonunifilar presentation of SNS Process:



$$\mathcal{B} = \{0, 1\}$$

$$T^{(0)} = \begin{pmatrix} 0 & 0 \\ \frac{1}{2} & 0 \end{pmatrix} \quad T^{(1)} = \begin{pmatrix} \frac{1}{2} & \frac{1}{2} \\ 0 & \frac{1}{2} \end{pmatrix}$$

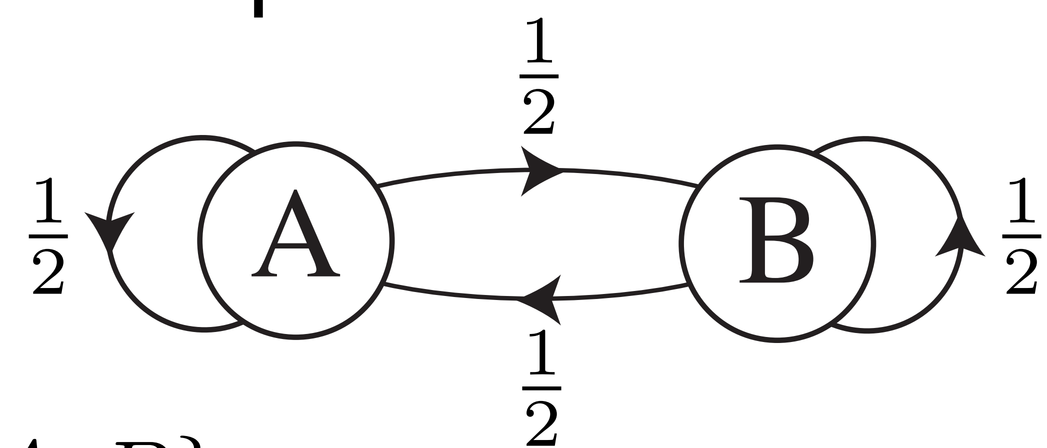
Measures of Complexity ...

Measures from the $\epsilon\mathrm{M}$:

Method A:

Calculate entropy rate from nonunifilar presentation:

Internal Markov chain:



Internal (= Fair Coin): $V = \{A, B\}$

$$T = \begin{pmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} \end{pmatrix} \quad \pi_V = \left(\tfrac{1}{2}, \tfrac{1}{2}\right)$$

Entropy rate?

$$\text{“}h_\mu^A(\mathrm{SNS})\text{”} = -\sum_{v \in V} \Pr(v) \sum_{v' \in V} \Pr(v'|v) \log_2 \Pr(v'|v)$$

$= 1$ bit/symbol

Measures of Complexity ...

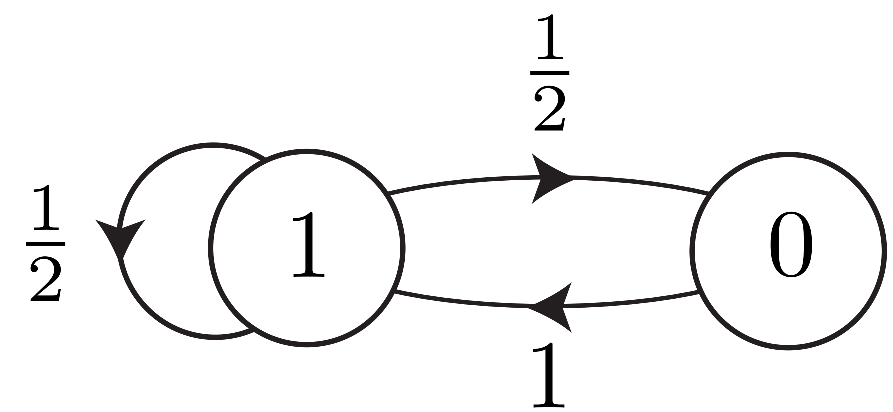
Measures from the $\epsilon\mathrm{M}$:

Method B:

Lower bound by entropy rate for "support" process:

No consecutive 0s!

Support process is GMP



$$T = \begin{pmatrix} \frac{1}{2} & \frac{1}{2} \\ 1 & 0 \end{pmatrix} \quad \pi_V = (2/3, 1/3)$$

A restriction that lowers entropy rate:

$$h_\mu^B(\mathrm{GMP}) = -\tfrac{2}{3} \sum_{s \in \mathcal{B}} \Pr(s|1) \log_2 \Pr(s|1) + 0$$

$$= \tfrac{2}{3} \approx 0.666667 \text{ bits/symbol}$$

Measures of Complexity ...

Measures from the $\epsilon\mathrm{M}$:

"SNS entropy rate" larger than support process?

$$\text{“}h_\mu^A(\mathrm{SNS})\text{”} \gg h_\mu^B(\mathrm{GMP})$$

Not possible!

Entropy of distribution always $\leq$ Entropy of support

Use of SNS presentation leads to overestimate of entropy rate: internal process more random than observed process.

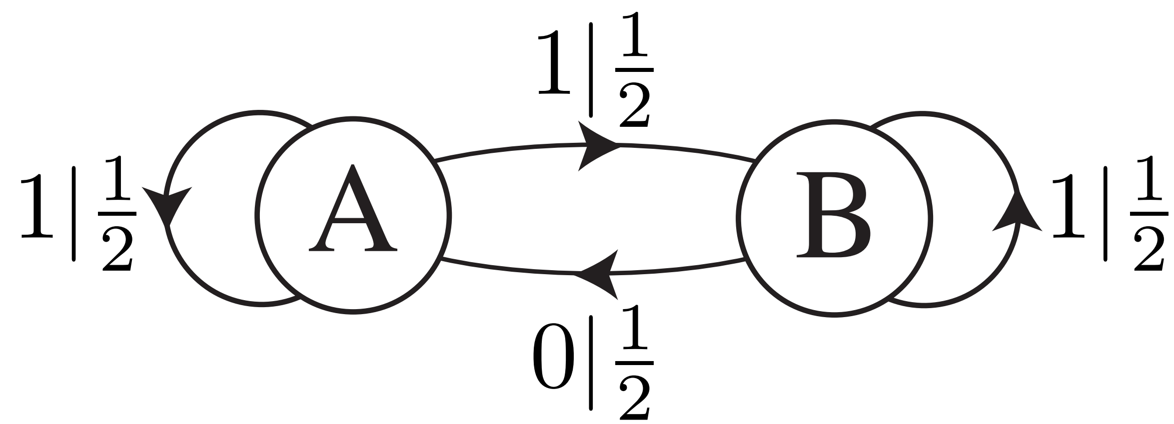
So how to compute SNS process's entropy rate?

Measures of Complexity ...

Measures from the ϵM:

Method C:

Estimate entropy rate from output process:



$$\text{``}h_\mu^C(\text{SNS})\text{''} = -\sum_{v \in V} \Pr(v) \sum_{(s,v') \in (\mathcal{A},V)} \Pr(s,v'|v) \log_2 \Pr(s,v'|v)$$

$$= -\left(\tfrac{1}{2} \times 0 - \tfrac{1}{2} \times 1\right)$$

$$= \tfrac{1}{2} \text{ bits/symbol}$$

Measures of Complexity ...

Measures from the $\epsilon\mathrm{M}$:

Entropy rate?

$$\text{“}h_\mu^A(\mathrm{SNS})\text{”} \gg h_\mu^B(\mathrm{GMP}) \gg \text{“}h_\mu^C(\mathrm{SNS})\text{”}$$

Lessons:

1. Cannot use nonunifilar HMM presentations for entropy rate.
2. Need $\epsilon\mathrm{M}$ to calculate entropy rate.

SNS example: Nontrivial, countably infinite $\epsilon\mathrm{M}$.

$$h_\mu(\mathrm{SNS}) \approx 0.6778 \text{ bits/symbol}$$

Curious:

Even to estimate a process's intrinsic randomness, need to infer its structure.

Measures of Complexity ...

Measures from the $\epsilon\mathrm{M}$...

Statistical Complexity of $\epsilon\mathrm{M}$:

$$C_\mu(\boldsymbol{\mathcal{S}}) = - \sum_{\mathcal{S} \in \boldsymbol{\mathcal{S}}} \Pr(\mathcal{S}) \log_2 \Pr(\mathcal{S})$$

where $\Pr(\mathcal{S})$ is casual-state asymptotic probability.

Meaning:

Shannon information in the causal states.

Measures of Complexity ...

Measures from the ϵM...

Statistical Complexity of a Process:

$$C_\mu \left(\Pr(\overleftrightarrow{S}) \right) = C_\mu(\boldsymbol{\mathcal{S}})$$

Meaning:

The amount of historical information a process stores.

The amount of structure in a process.

Measures of Complexity ...

Measures from the $\epsilon\mathrm{M}$...

Excess Entropy: Three versions, all equivalent for 1D processes

$$\mathbf{E} = \lim_{L \to \infty} [H(L) - h_\mu L]$$

$$\mathbf{E} = \sum_{L=1}^{\infty} [h_\mu(L) - h_\mu]$$

$$\mathbf{E} = I[\overleftarrow{S}; \overrightarrow{S}]$$

How to get, given $\epsilon\mathrm{M}$?

Special cases: When $\epsilon\mathrm{M}$ is IID, periodic, or spin chain.

General case: Need a new framework ... next lectures.

Measures of Complexity ...

Measures from the $\epsilon\mathrm{M}$...

Excess Entropy ...

$\epsilon\mathrm{M}$ is IID:

$$\mathbf{E} = 0 \qquad C_\mu = 0$$

$\epsilon\mathrm{M}$ is Period P:

$$\mathbf{E} = \log_2 P = C_\mu$$

$\epsilon\mathrm{M}$ is range-R spin chain:

$$\mathbf{E} = H(R) - R h_\mu$$

$$\mathbf{E} = C_\mu - R h_\mu$$

Measures of Complexity ...

Measures from the $\epsilon\mathrm{M}$...

Excess Entropy ...

Typically for Markov Chains:

$$\mathbf{E} < C_\mu$$

What can be said in general?

Measures of Complexity ...

Measures from the $\epsilon\mathrm{M}$ ...

Bound on Excess Entropy:

$$\mathbf{E} \leq C_\mu$$

Proof sketch:

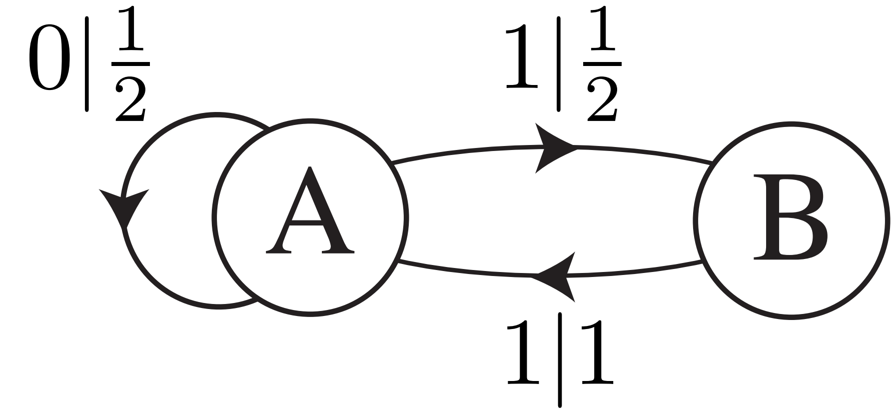
(1) $\mathbf{E} = I[\overrightarrow{S};\overleftarrow{S}] = H[\overrightarrow{S}] - H[\overrightarrow{S} \mid \overleftarrow{S}]$

(2) Causal States: $H[\overrightarrow{S} \mid \overleftarrow{S}] = H[\overrightarrow{S} \mid \mathcal{S}]$

(3)
$$\begin{aligned}
\mathbf{E} &= H[\overrightarrow{S}] - H[\overrightarrow{S} \mid \mathcal{S}] \\
&= I[\overrightarrow{S};\mathcal{S}] \\
&= H[\mathcal{S}] - H[\mathcal{S} \mid \overrightarrow{S}] \\
&\leq H[\mathcal{S}] = C_\mu
\end{aligned}$$

Measures of Complexity ...

Measures from the $\epsilon\mathrm{M}$ ...



Bound on Excess Entropy ...

But, the bound is saturated!

Even process:

$$T^{(0)} = \begin{pmatrix} \frac{1}{2} & 0 \\ 0 & 0 \end{pmatrix}$$

$$T^{(1)} = \begin{pmatrix} 0 & \frac{1}{2} \\ 1 & 0 \end{pmatrix}$$

$$\pi_V = (2/3, 1/3)$$

$$C_\mu = H(2/3) \approx 0.9182$$

$$\mathbf{E} \approx 0.9182$$

When does this occur?

In general, need a new framework for answering this question.

Measures of Complexity ...

Measures from the $\epsilon\mathrm{M}$ ...

Bound on Excess Entropy ...

Consequence: The Cryptographic Limit

Can have $\mathbf{E} \rightarrow 0$ when $C_\mu \gg 1$.

State information is still there,
just hidden in the observed sequences.

Goal of cryptography:
Cyphertext appears random to all observers,
but still contains the original information,
recoverable by the intended recipient.

Measures of Complexity ...

Measures from the ϵM ...

Bound on Excess Entropy ...

For example: The Almost IID Processes

$$p_i \sim \frac{1}{2} + \ \mathrm{Uniform}(0, \epsilon)$$

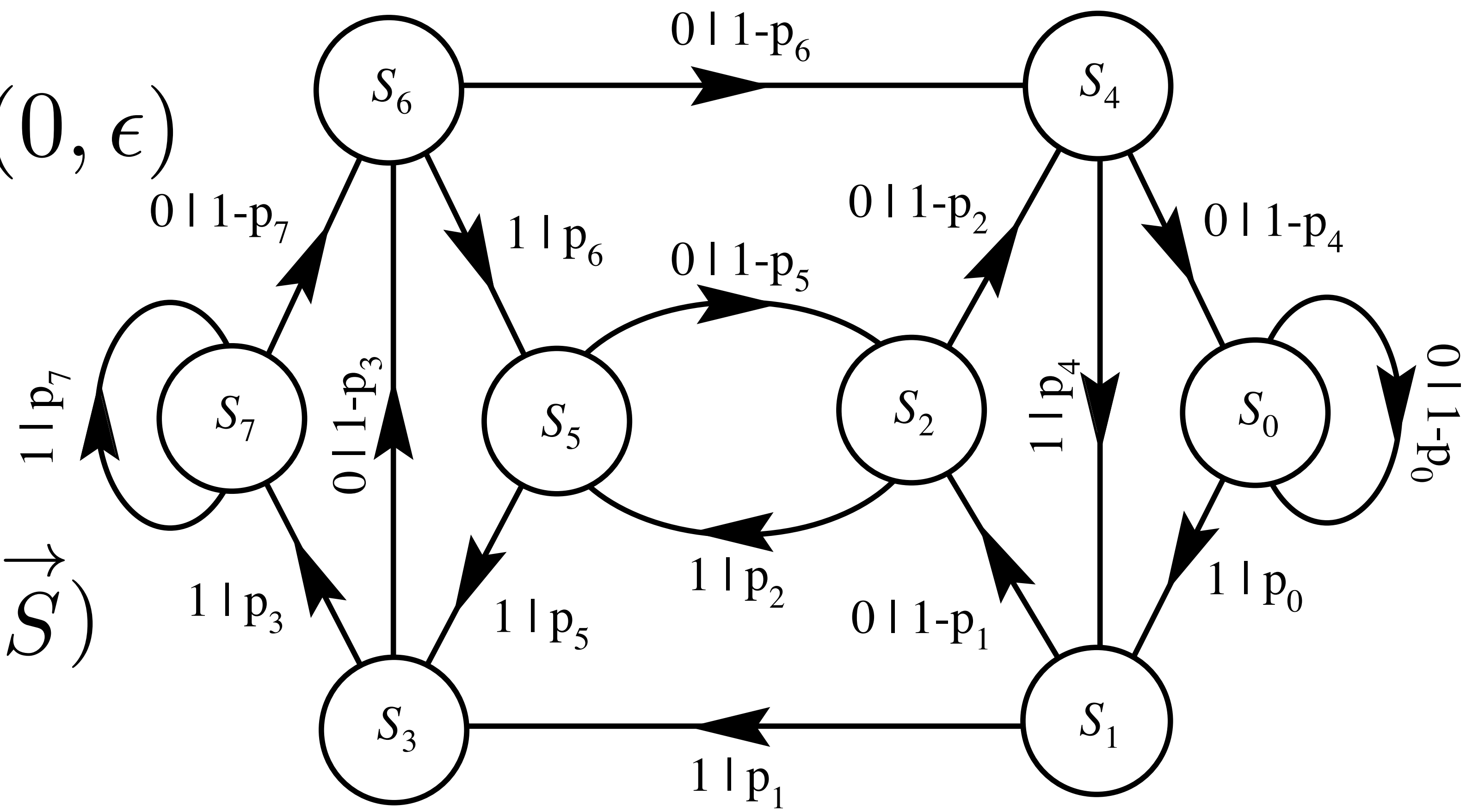
$$\epsilon \searrow 0$$

$$\Pr(\overleftrightarrow{S}) \approx \Pr(\overleftarrow{S})\Pr(\overrightarrow{S})$$

$$\mathbf{E} \approx 0$$

$$\Pr(\mathcal{S}_i) \approx |\boldsymbol{\mathcal{S}}|^{-1}$$

$$C_\mu \approx \log_2 |\boldsymbol{\mathcal{S}}|$$



Measures of Complexity ...

Measures from the $\epsilon\mathrm{M}$ ...

Bound on Excess Entropy ... $\mathbf{E} \leq C_\mu$

Consequence:

Excess entropy is *not* the process' stored information.

Statistical complexity C_μ *is* the stored information.

$\mathbf{E}$ is the *apparent* information,
as revealed in *measurement sequences*.

Measures of Complexity ...

Measures from the ϵM ...

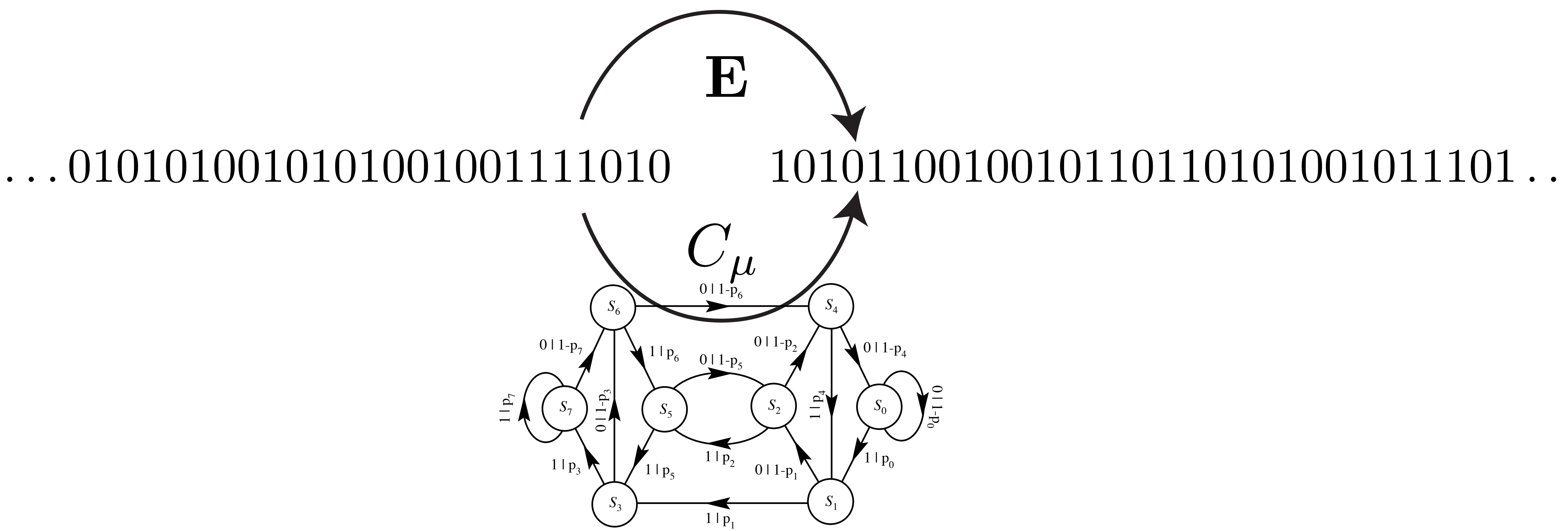
Bound on Excess Entropy ...

Executive Summary:

C_μ is the amount of information the process uses

to *communicate*

$\mathbf{E}$ bits of information from the past to the future.



Measures of Complexity ...

Measures from the $\epsilon\mathrm{M}$ ...

Bound on Excess Entropy: $\mathbf{E} \leq C_\mu$

Consequence:

The inequality is Why We Must Model.

Cannot simply use sequences as states.

There is internal structure not expressed by sequences.

Widespread misconception about excess entropy $I[\overleftarrow{X};\overrightarrow{X}]$:
"the predictive information".
No! C_μ is the information required to optimally predict.
Excess entropy is the "predictable" information:
that which can be predicted.

Measures of Complexity ...

Measures from the $\epsilon\mathrm{M}$ ...

Bound on Predictability (Redundancy):

$$r(1) \leq C_\mu$$

Proof sketch:

(1) $r(1) = H[\overrightarrow{S}^1] - h_\mu$

(2)
$$\begin{aligned}
H[\overrightarrow{S}^1] - h_\mu &= H[\overrightarrow{S}^1] - H[\overrightarrow{S}^1 \mid \overleftarrow{S}] \\
&= H[\overrightarrow{S}^1] - H[\overrightarrow{S}^1 | \mathcal{S}] \\
&= I[\overrightarrow{S}^1 ; \mathcal{S}] \\
&\leq H[\mathcal{S}] \\
&= C_\mu
\end{aligned}$$

Measures of Complexity ...

An $\epsilon\mathrm{M}$ Captures "Pattern":

Symmetry = group:

Rotate square 90° in plane.
Square appears in original orientation.

Square "shape" is invariant under 90° rotation.

Inverse: Rotate by -90°.

An $\epsilon\mathrm{M}$ Captures "Pattern" ...

$\epsilon\mathrm{M}$ is a monoid:

$\{T_{ij}^{(s)}\}$ form a semi-group with identity.
Semi-group is a group without a unique inverse.

When semi-group is a group: $\epsilon\mathrm{M}$ algebra describes a symmetry.

Semi-groups describe "generalized", noisy, wild-card symmetries.

An $\epsilon\mathrm{M}$ Captures "Pattern":

Example of Group: Period-2 process

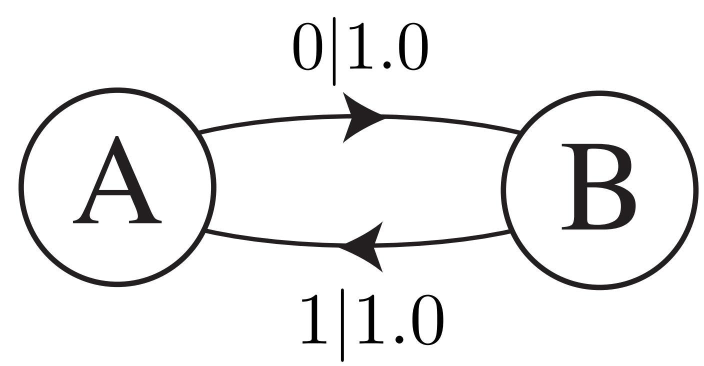
$$\ldots 010101010101010101010101 \ldots$$

$$\ldots 010101010101010101010101 \ldots$$

Temporal translation symmetry:

Shift time by two steps ($t \rightarrow t+2$), get same sequence.

$\epsilon\mathrm{M}$ period-2 in the causal states:



$$\ldots ABABABABABAB \ldots$$

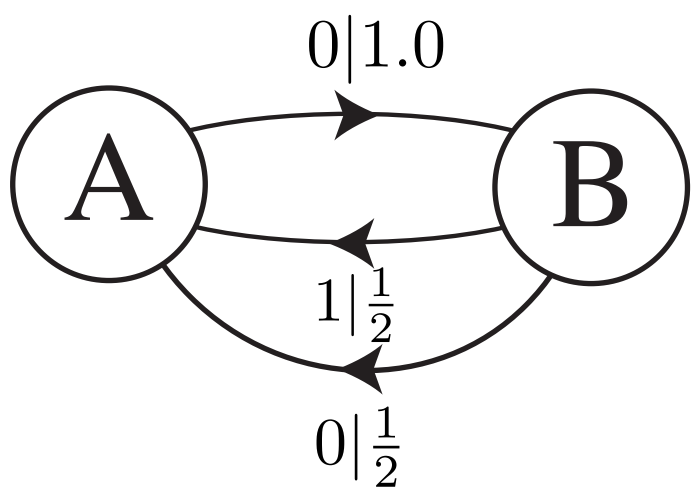
Measures of Complexity ...

An $\epsilon\mathrm{M}$ Captures "Pattern":

Example: Random process

$$\ldots 0101000100010101000010001 \ldots$$

$\epsilon\mathrm{M}$:



Period-2 in the causal states: Detects the order in chaos

$$\ldots ABABABABABAB \ldots$$

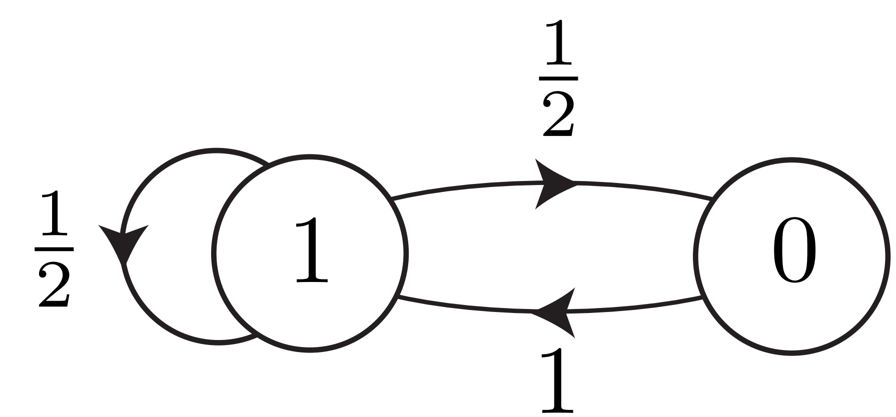
Temporal translation symmetry:

Shift time by two steps ($t \to t+2$), get same statistics.

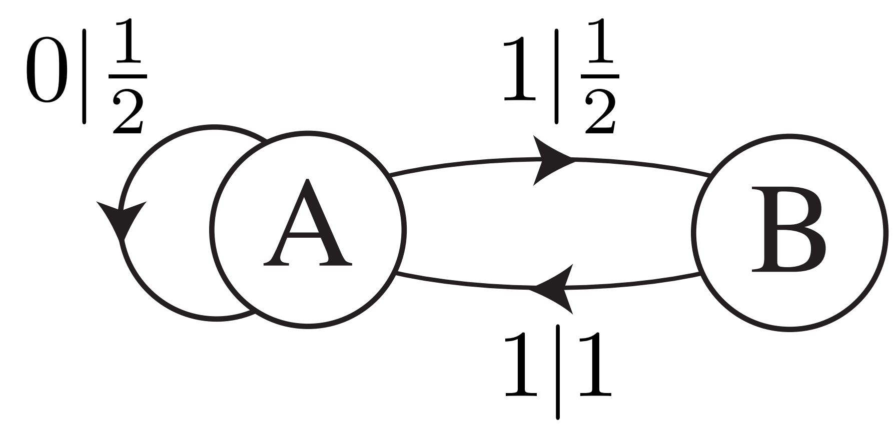
Measures of Complexity ...

An $\epsilon\mathrm{M}$ Captures "Pattern":

Example: GMS "pattern" = No consecutive 0s
By inspecting $\epsilon\mathrm{M}$ structure.



Example: Even process "pattern" = "Evenness"
By inspecting $\epsilon\mathrm{M}$ structure.

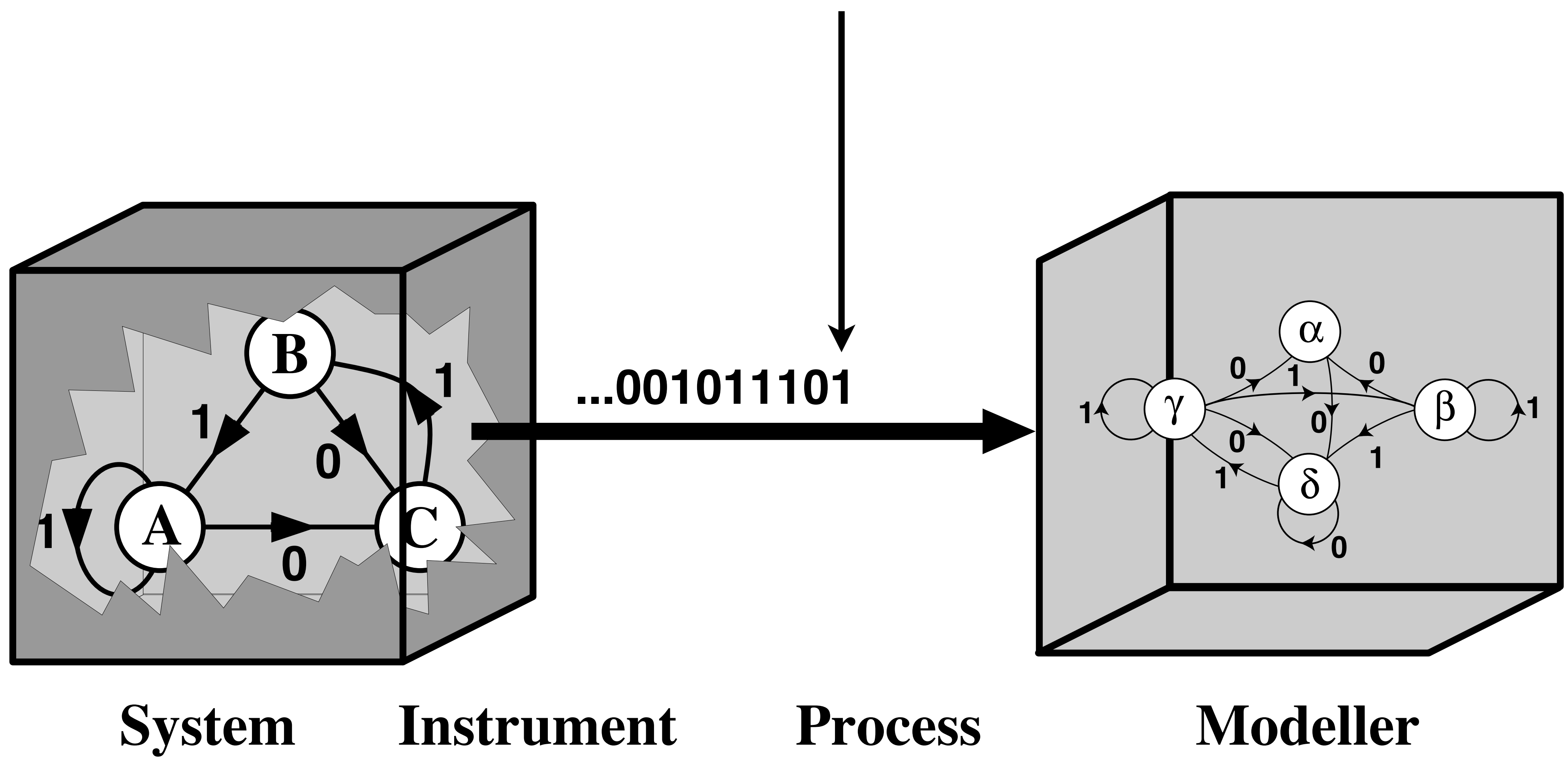


Measures of Complexity ...

An $\epsilon\mathrm{M}$ Captures "Pattern":

Measurement semantics:

What does a particular measurement mean?



Measurement Channel

Measures of Complexity ...

An ϵM Captures "Pattern":

Measurement semantics: Prediction level

What is the meaning of a particular measurement s?

Shannon says the amount of "information" in s is:

$$-\log_2 \Pr(\text{observing } s)$$

Given ϵM (assuming you're sync'd to internal (causal) state):

$$-\log_2 \Pr(\text{observing } s) = -\log_2 \Pr(\mathcal{S} \rightarrow_s \mathcal{S}')$$

Measures of Complexity ...

An $\epsilon\mathrm{M}$ Captures "Pattern":

Measurement semantics: Prediction level ...

Example: $t = 0\ 1\ 2\ 3\ 4\ 5\ 6\ 7\ 8\ 9\ 10\ 11$

$s = 0\ 1\ 1\ 1\ 1\ 0\ 1\ 1\ 0\ 1\ \ 1\ \ 1$

At $t = 11$ measure $s_{11} = 1$

How much information does this give?

$$H(s_{11}|s_{10} = 1, s_9 = 1, \ldots) \approx h_\mu (\approx 0.585 \text{ bits})$$

Degree of observer's surprise (predictability)

Does not say what the event $s_{11} = 1$ means to the observer!

Measures of Complexity ...

An $\epsilon\mathrm{M}$ Captures "Pattern":

Measurement semantics ...

Meaning: Tension between representations of same event at different levels; e.g.:

Level 1 is data stream and the event is a measurement
Level 2 is the agent and the event updates it's model

Degree of meaning of observing $s \in \mathcal{A}$:

$$\Theta(s) = -\log_2 \Pr(\rightarrow_s \mathcal{S})$$

where $\mathcal{S}$ is the causal state to which s brings observer.

Meaning content: State selected from anticipated palette.

Measures of Complexity ...

An $\epsilon\mathrm{M}$ Captures "Pattern":

Measurement semantics ...

Meaningless: Start state (all futures possible)

$$\Theta(s) = -\log_2 \Pr(\mathcal{S}_0) = -\log_2 1 = 0 \qquad\qquad s = \lambda$$

Action on disallowed transition:

Reset to state of total ignorance (start state)

Disallowed transition is meaningless.

Meaningless measurements are informative, though:

$$-\log_2 \Pr(\mathcal{S} \rightarrow_s \mathcal{S}_0) = -\log_2 0 = \infty$$

Measures of Complexity ...

An $\epsilon\mathrm{M}$ Captures "Pattern":

Measurement semantics ...

Theorem:

$$\langle \Theta(s) \rangle = C_\mu$$

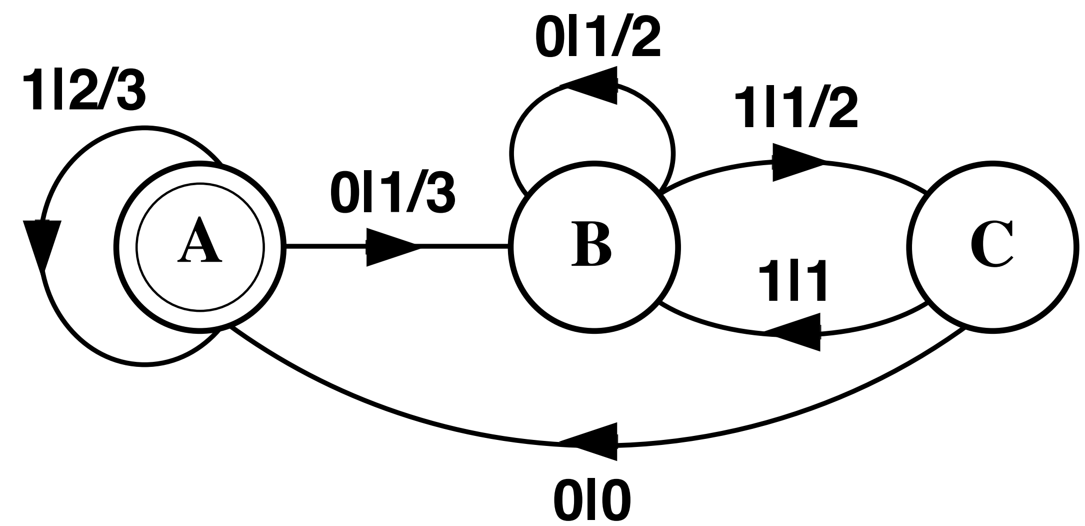
Average amount of meaning is the Statistical Complexity.

Measures of Complexity ...

An ϵM Captures "Pattern":

Measurement semantics: Example

...110110111011111101101...



Observer's Semantic Analysis				
State	Measurement	Surprise (bits)	Semantic State: Meaning	Degree of Meaning [bits]
A	λ	Not defined	No measurement	0
A	1	0.585	A: Unsynchronized	0.585 ... oo
A	0	1.585	B: Synchronized	1.585 ... 0.585
B	1	1	C: Odd # 1s	1.585
B	0	1	B: Even # 1s	0.585
C	1	0	B: Even # 1s	0.585
C	0	∞	A: Loose sync; reset	0

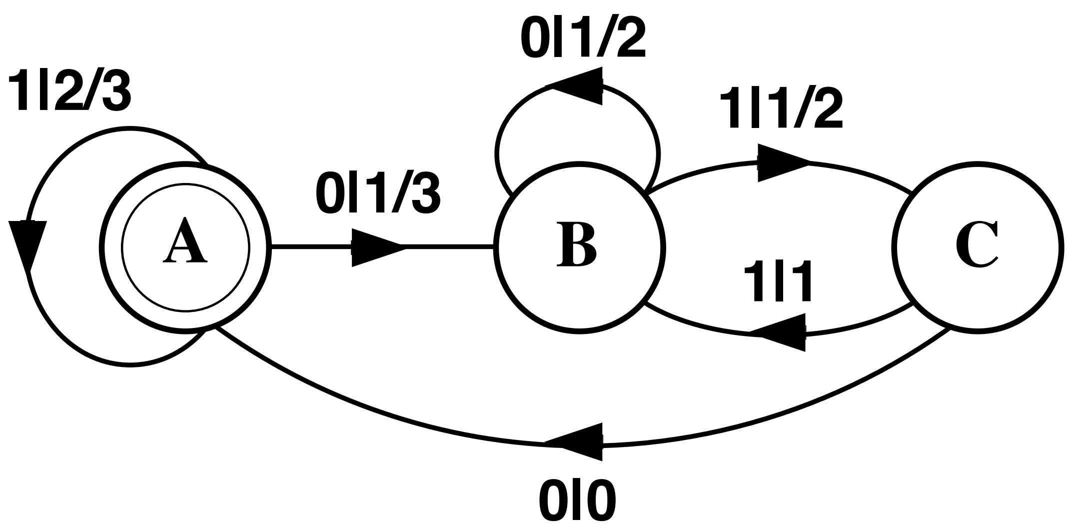
... the Even Process

Measures of Complexity ...

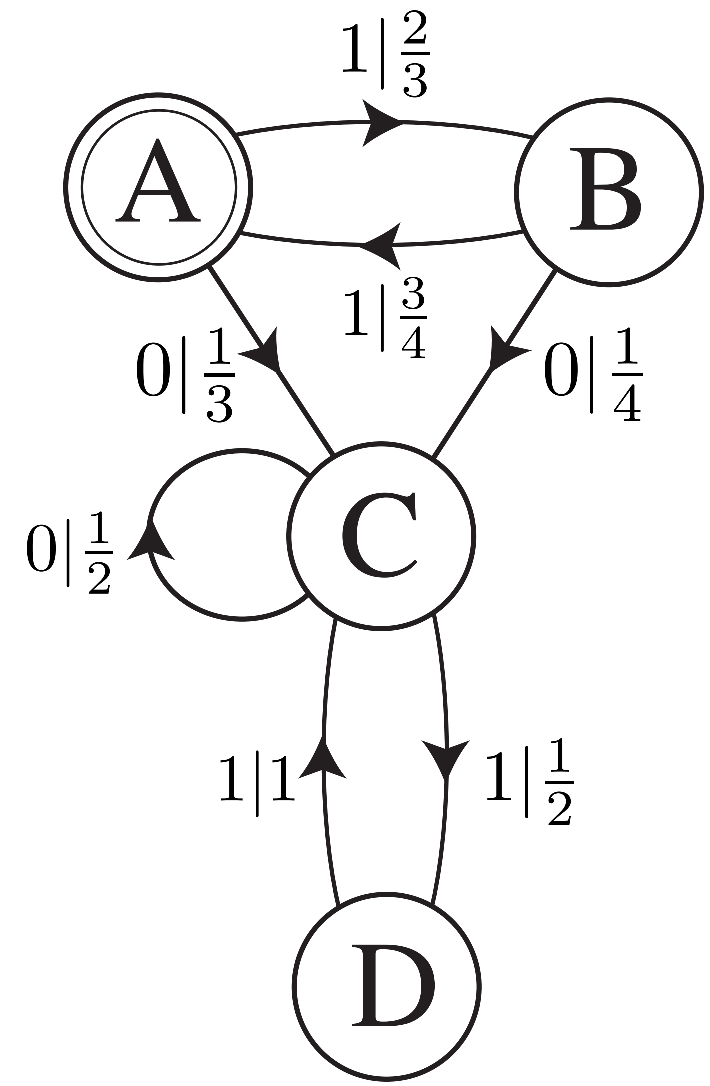
An $\epsilon\mathrm{M}$ Captures "Pattern":

Measurement semantics:

... the Even Process?



No, not exactly:



Semantics induced by any model:
Even an incorrect one!

Measures of Complexity ...

An $\epsilon\mathrm{M}$ Captures "Pattern":

Measurement semantics: Slipped something by ...

Preceding was about the *amount* of meaning!

Related to amount of memory/structure

But what is the meaning content?

Meaning = Algebra of the $\epsilon\mathrm{M}$

Measures of Complexity ...

An $\epsilon\mathbb{M}$ Captures "Pattern":

Measurement semantics ...

D. M. MacKay, "Information, Meaning & Mechanism", MIT (1969):

Meaning as selection of anticipated context.

Here:

$\epsilon\mathbb{M}$s give the natural set of contexts (causal states)
for an information source.

Measures of Complexity ...

Reading for next lecture:

CMR articles TBA and PRATISP.