Information in Processes

Entropy Growth for Stationary Stochastic Processes: $\Pr(\overleftrightarrow{S})$

Block Entropy:

$$H(L) = H(\Pr(s^L)) = -\sum_{s^L \in \mathcal{A}} \Pr(s^L) \log_2 \Pr(s^L)$$

Monotonic increasing: $H(L) \geq H(L-1)$

Adding a random variable cannot decrease entropy:

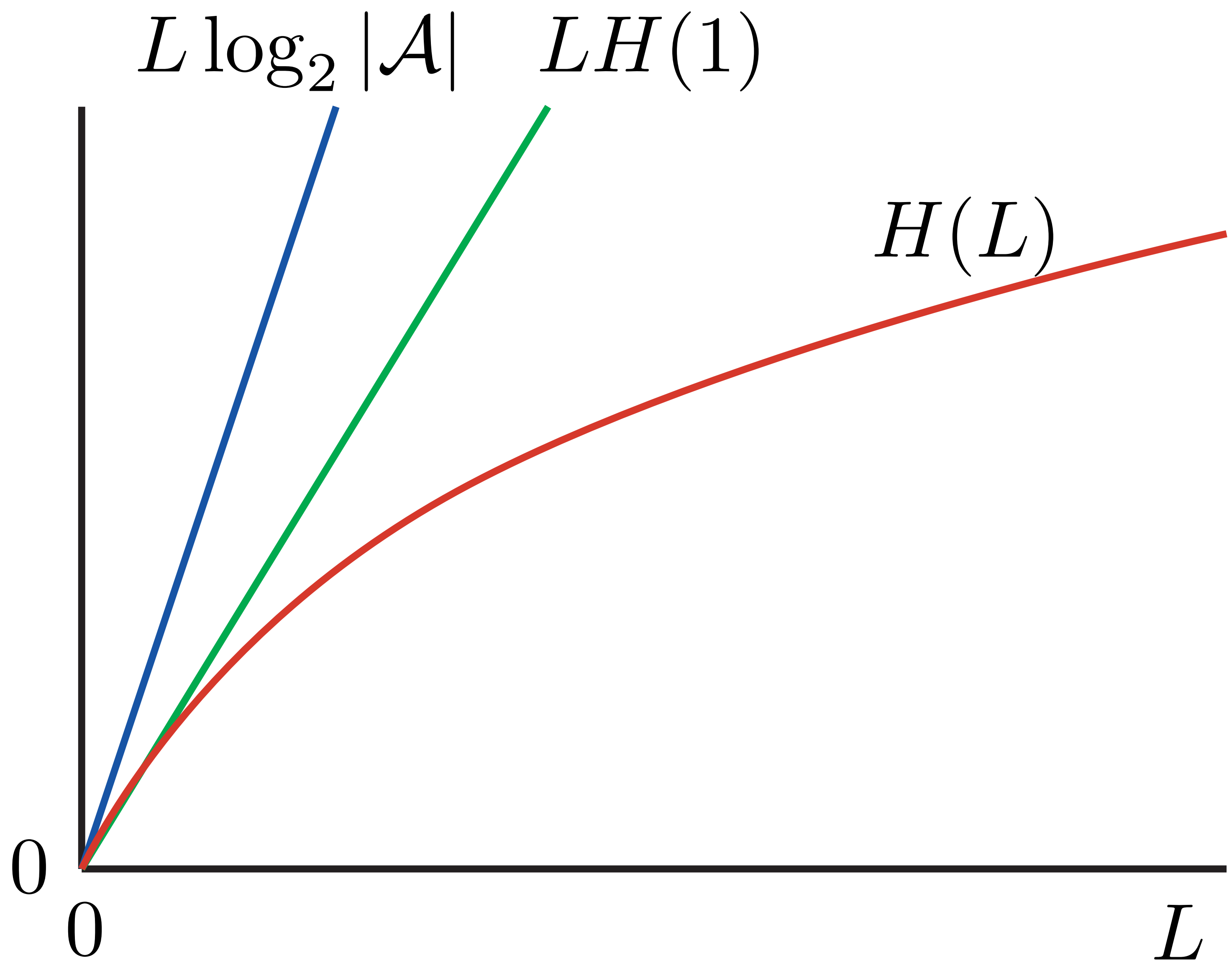
$$H(S_1, S_2, \ldots, S_L) \leq H(S_1, S_2, \ldots, S_L, S_{L+1})$$

No measurements, no information: $H(0) = 0$

Information in Processes ...

Entropy Growth for Stationary Stochastic Processes ...

Block Entropy ...



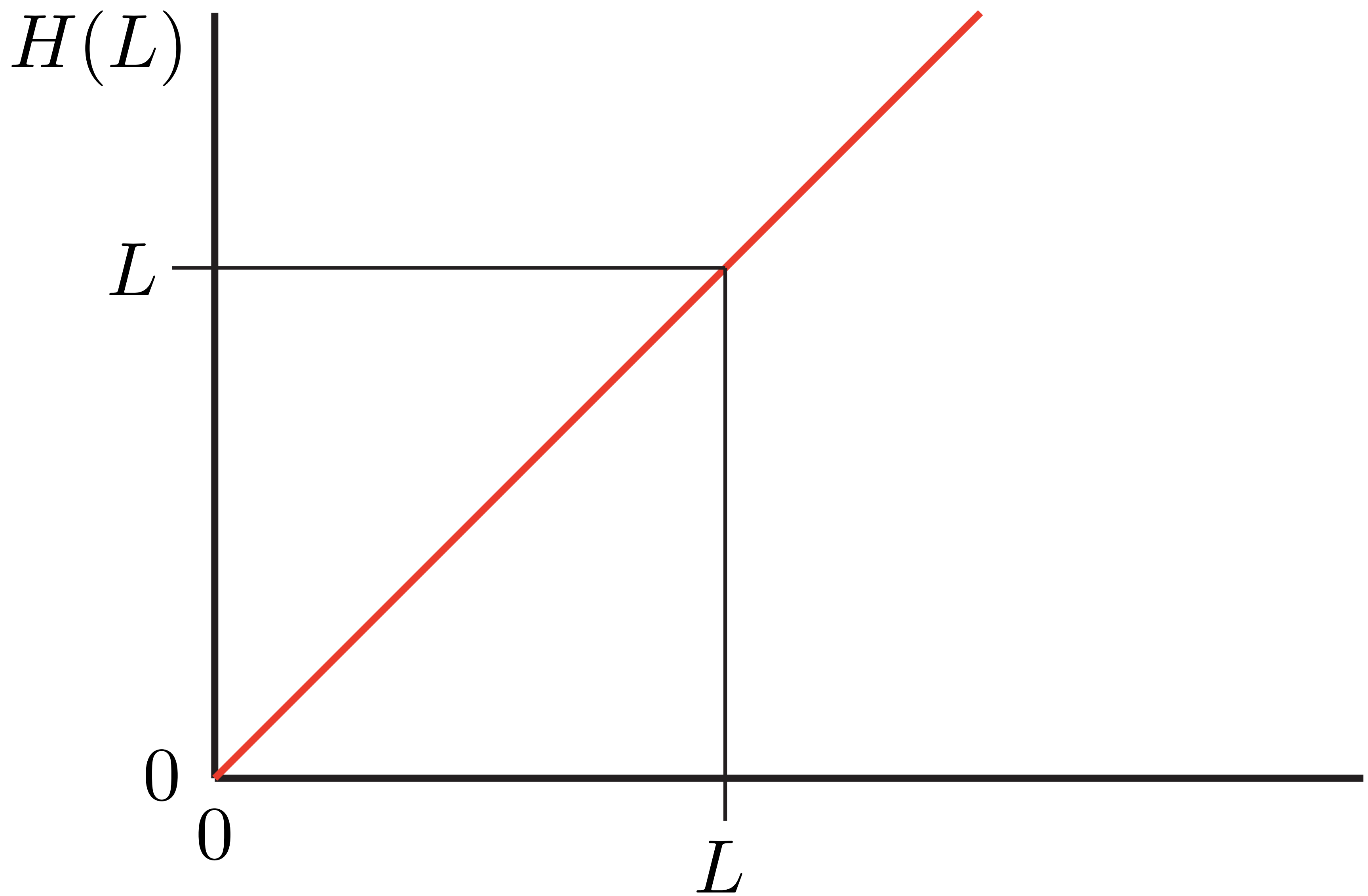
Entropy Growth for Stationary Stochastic Processes ...

Block Entropy ...

Example: Fair Coin

$$\Pr(s^L) = \frac{1}{2^L}$$

$$H(L) = L$$



Information in Processes ...

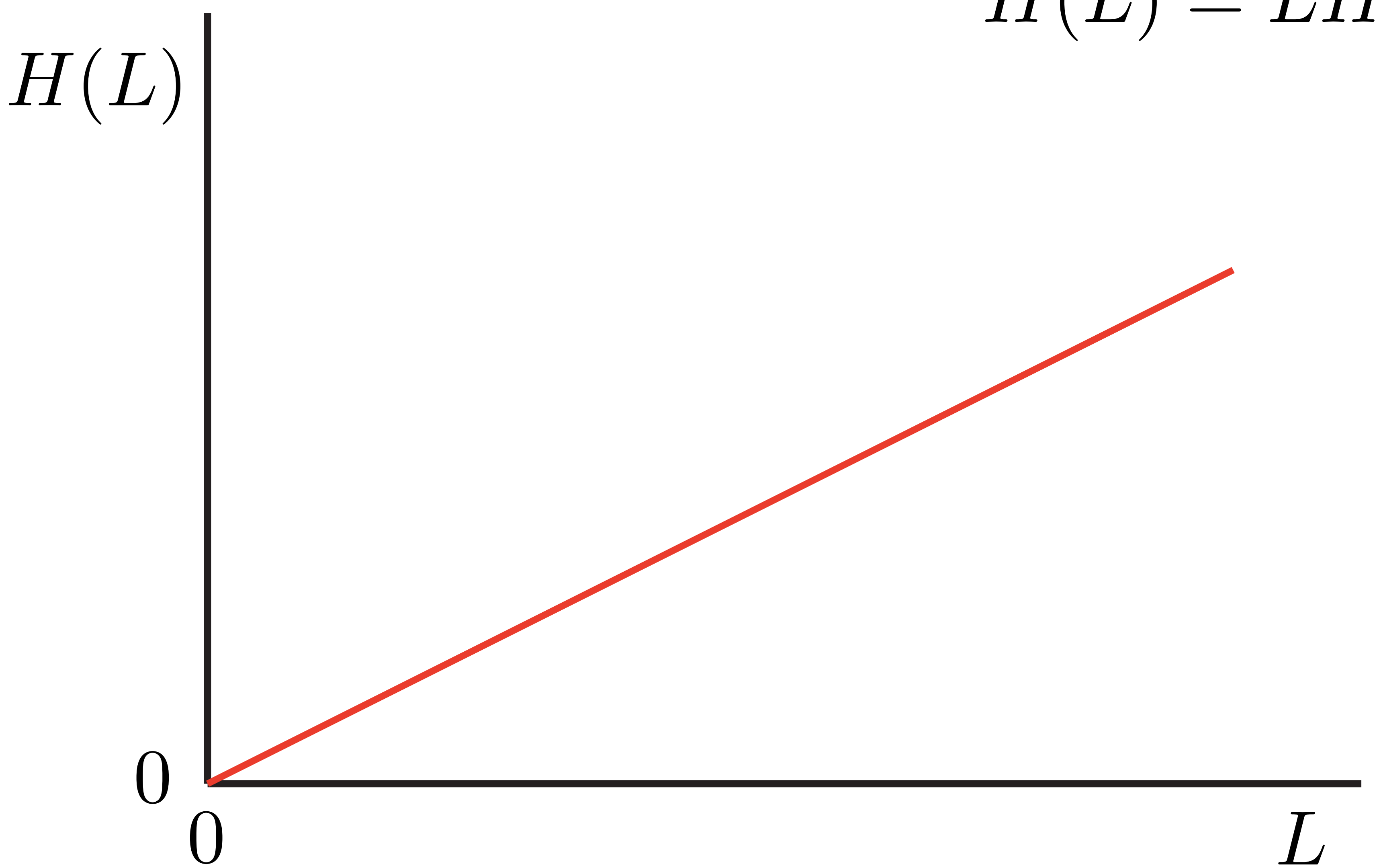
Entropy Growth for Stationary Stochastic Processes ...

Block Entropy ...

Example: Biased Coin

$$\Pr(s^L) = p^n (1-p)^{L-n}$$

$$H(L) = L H(p)$$



For any IID process:

$$H(L) = L H(S_1)$$

Information in Processes ...

Entropy Growth for Stationary Stochastic Processes ...

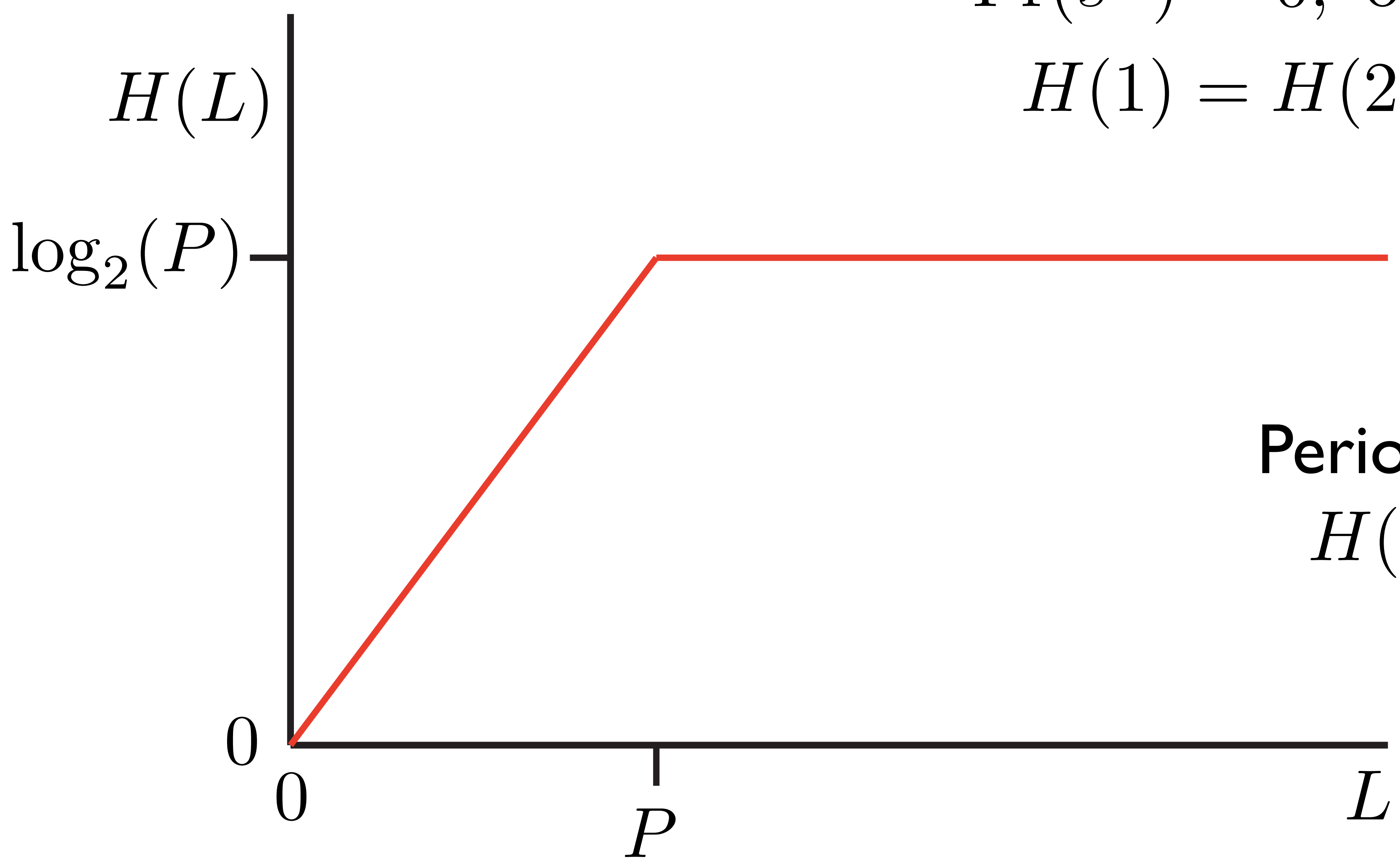
Block Entropy ...

Example: Period-2 Process

$$\ldots 0101010101 \ldots$$

$$\Pr(0) = \Pr(1) = \frac{1}{2}$$
$$\Pr(01) = \Pr(10) = \frac{1}{2}$$
$$\Pr(101) = \Pr(010) = \frac{1}{2}$$
$$\vdots$$
$$\Pr(s^L) = 0, \text{ otherwise}$$

$$H(1) = H(2) = H(L \geq 1) = 1$$



Period-P Process:

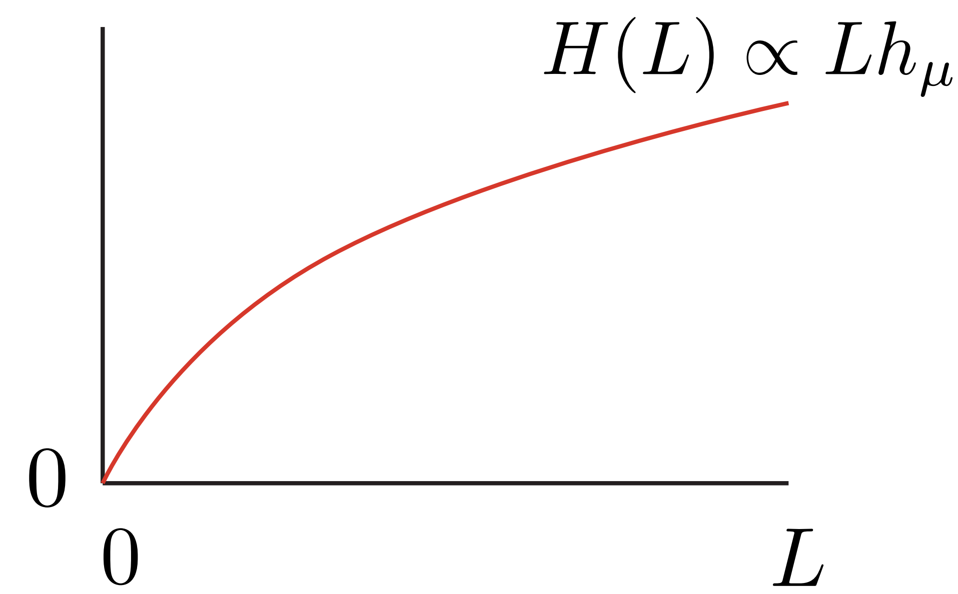
$$H(L \geq P) = \log_2(P)$$

Entropy Rates for Stationary Stochastic Processes:

Entropy per symbol is given by the Source Entropy Rate:

$$h_\mu = \lim_{L \rightarrow \infty} \frac{H(L)}{L}$$

(When limits exists.)



Interpretations:

- Asymptotic growth rate of entropy
- Irreducible randomness of process
- Average description length (per symbol) of process

Use: Typical sequences have probability: $\Pr(s^L) \approx 2^{-L \cdot h_\mu}$

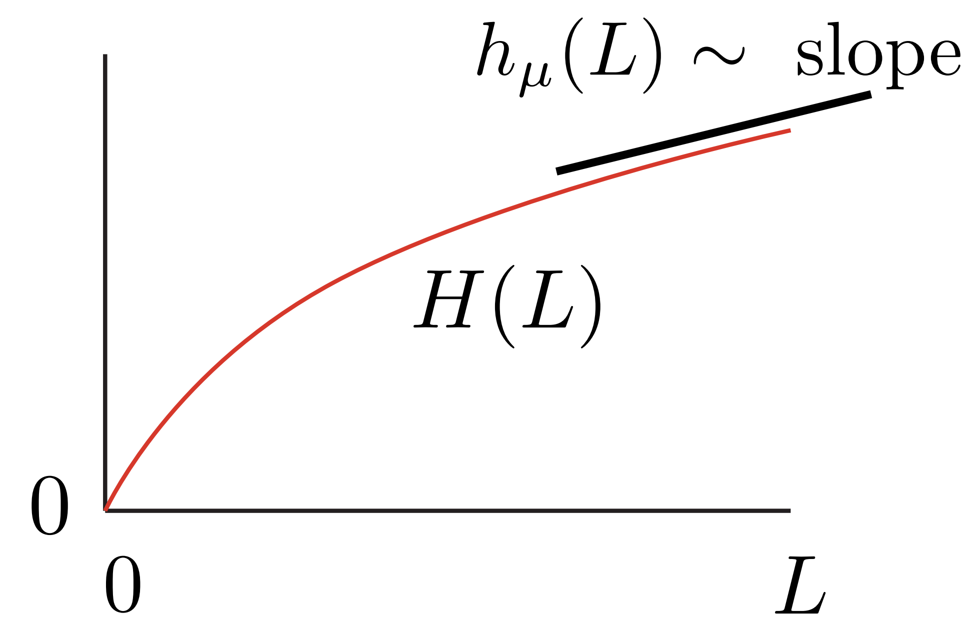
(Shannon-MacMillian-Breiman Theorem)

Entropy Rates for Stationary Stochastic Processes ...

Length-L Estimate of Entropy Rate:

$$\widehat{h}_\mu(L) = H(L) - H(L-1)$$

$$\widehat{h}_\mu(L) = H(s_L | s_1 \cdots s_{L-1})$$



Boundary condition:

$\widehat{h}_\mu(0) = \log_2 |\mathcal{A}|$: no measurements, all things possible

$$\widehat{h}_\mu(1) = H(1)$$

Monotonic decreasing: $\widehat{h}_\mu(L) \leq \widehat{h}_\mu(L-1)$

Conditioning cannot increase entropy:

$$H(s_L | s_1 \cdots s_{L-1}) \leq H(s_L | s_2 \cdots s_{L-1}) = H(s_{L-1} | s_1 \cdots s_{L-2})$$

Information in Processes ...

Entropy Rates for Stationary Stochastic Processes:

Entropy rate ...

$$\widehat{h}_\mu = \lim_{L \to \infty} \widehat{h}_\mu(L) = \lim_{L \to \infty} H(s_0 | \overleftarrow{s}^L) = H(s_0 | \overleftarrow{s})$$

Interpretations:

- Uncertainty in next measurement, given past
- A measure of unpredictability
- Asymptotic slope of block entropy

Alternate entropy rate definitions agree:

$$\widehat{h}_\mu = h_\mu$$

Entropy Rate for a Markov chain: $\{V, T\}$

$$
\begin{aligned}
h_\mu &= \lim_{L \to \infty} h_\mu(L) \\
&= \lim_{L \to \infty} H(v_L | v_1 \cdots v_{L-1}) \\
&= \lim_{L \to \infty} H(v_L | v_{L-1})
\end{aligned}
$$

Assuming asymptotic state distribution:

- Process in statistical equilibrium
- Process running for a long time
- Forgotten it's initial distribution

Closed-form:

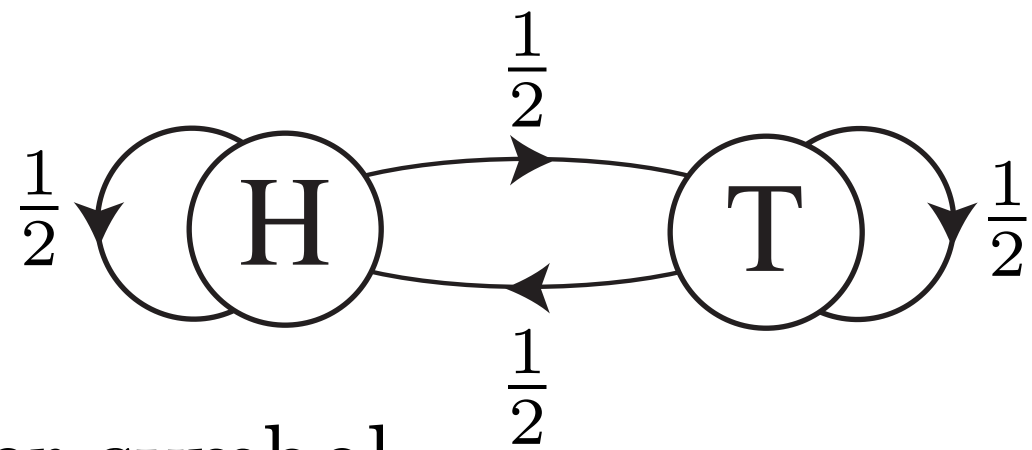
$$
h_\mu = - \sum_{v \in V} p_v(\infty) \sum_{v' \in V} T_{vv'} \log_2 T_{vv'}
$$

Information in Processes ...

Entropy Rate for Markov chains ...

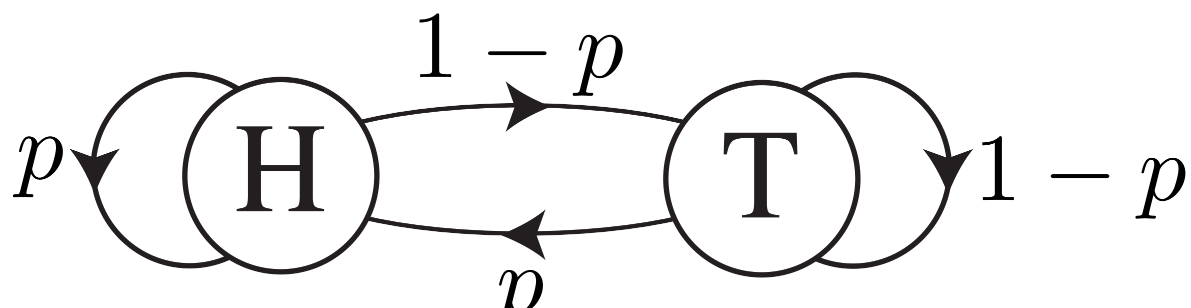
Examples:

(1) Fair Coin:



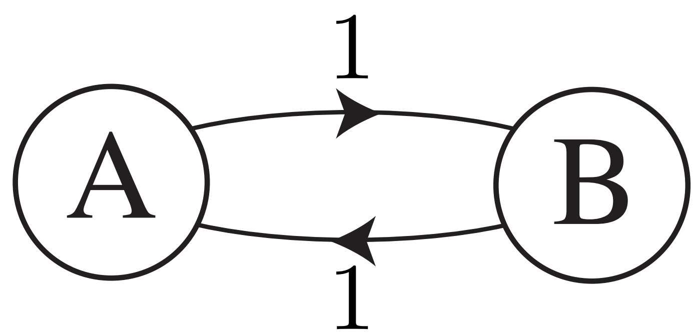
$$h_\mu = 1 \text{ bit per symbol}$$

(2) Biased Coin:



$$h_\mu = H(p) \text{ bits per symbol}$$

(3) Period-2 Process:



$$h_\mu = 0 \text{ bits per symbol}$$

Entropy Rate for Unifilar Hidden Markov Chain:

Internal: $\{V, T\}$
Observed: $\{T^{(s)} : s \in \mathcal{A}\}$

Closed-form for entropy rate:

$$h_\mu = - \sum_{v \in V} p_v(\infty) \sum_{s \in \mathcal{A}} \sum_{v' \in V} T^{(s)}_{vv'} \log_2 T^{(s)}_{vv'}$$

Due to unifilarity:
Observed sequences are (effectively) unique paths in UHMC

Entropy Rate for Unifilar Hidden Markov Chain ...

Example: Why are modems noisy?

Recall previous prefix code example

Distribution:
$$p(a) = \frac{1}{2}$$
$$p(b) = \frac{1}{4}$$
$$p(c) = \frac{1}{8}$$
$$p(d) = \frac{1}{8}$$

$$H(X) = 1.75 \text{ bits}$$

Codebook:
$$C(a) = 0$$
$$C(b) = 10$$
$$C(c) = 110$$
$$C(d) = 111$$

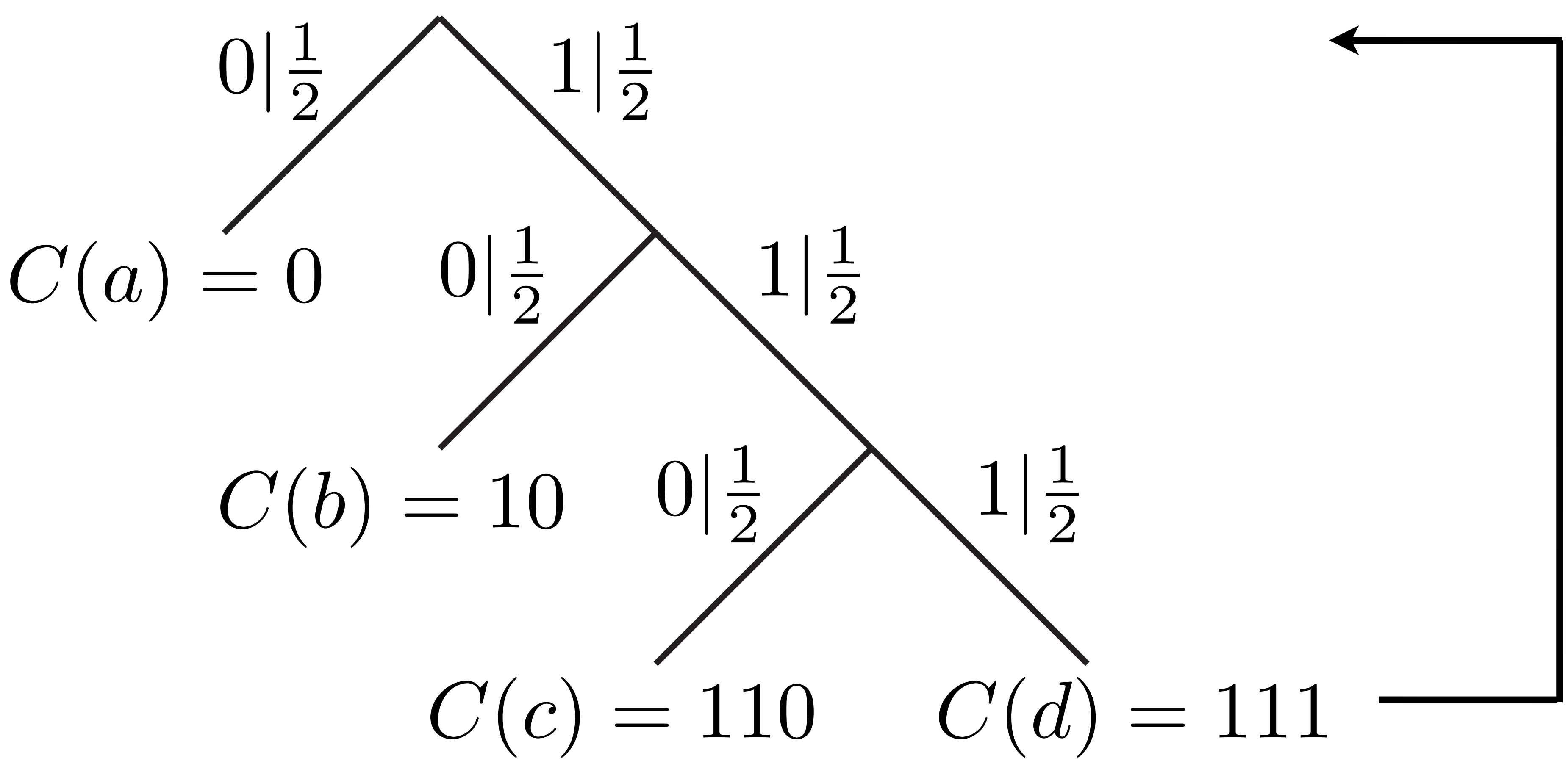
$$R(C) = 1.75 \text{ bits per message}$$

What is entropy rate (per output bit) of encoded stream?

Entropy Rate for Deterministic Hidden Markov Chain ...

Example: Why are modems noisy?

How often are codewords generated?



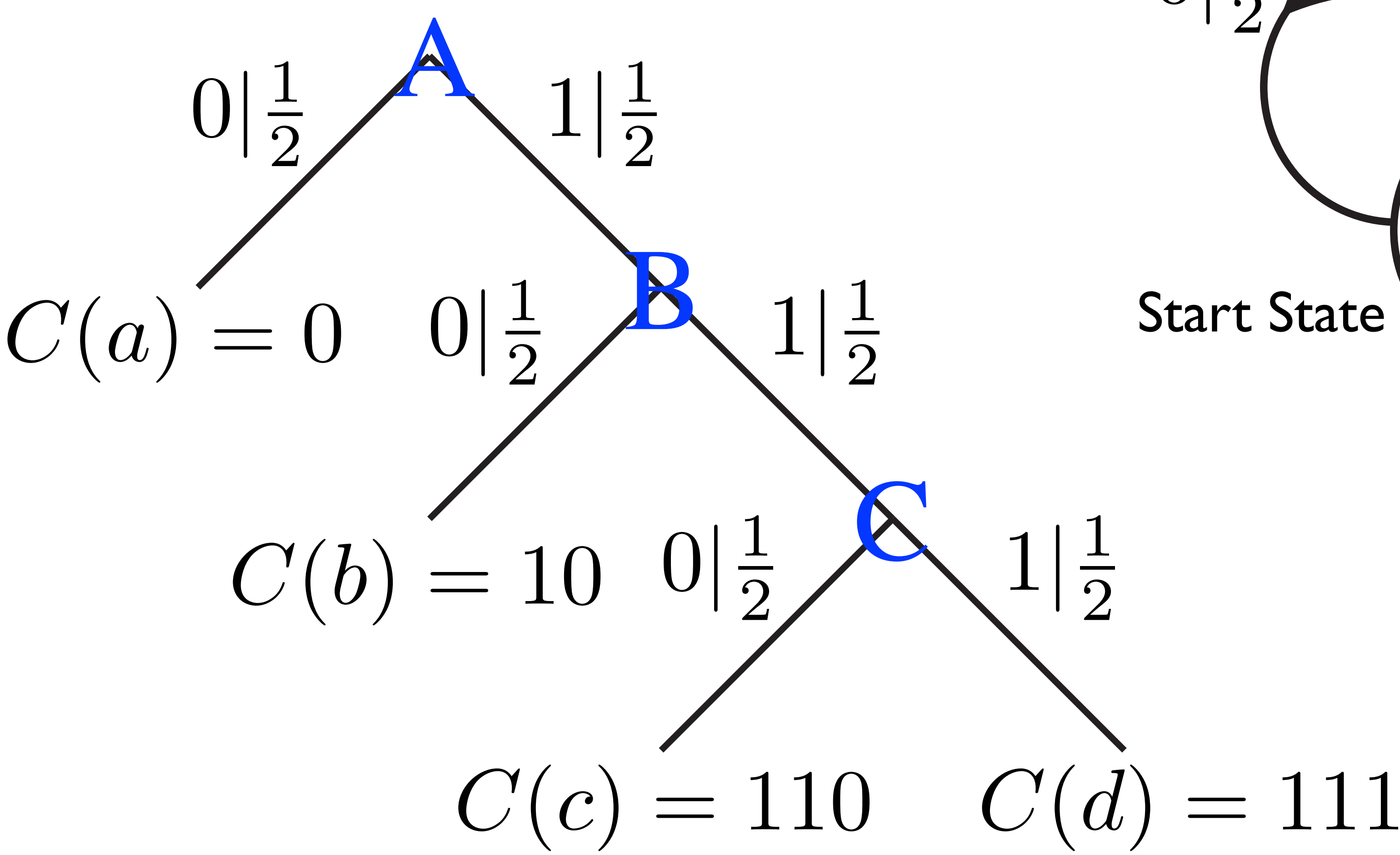
Encoding (output of channel) is a hidden Markov chain:

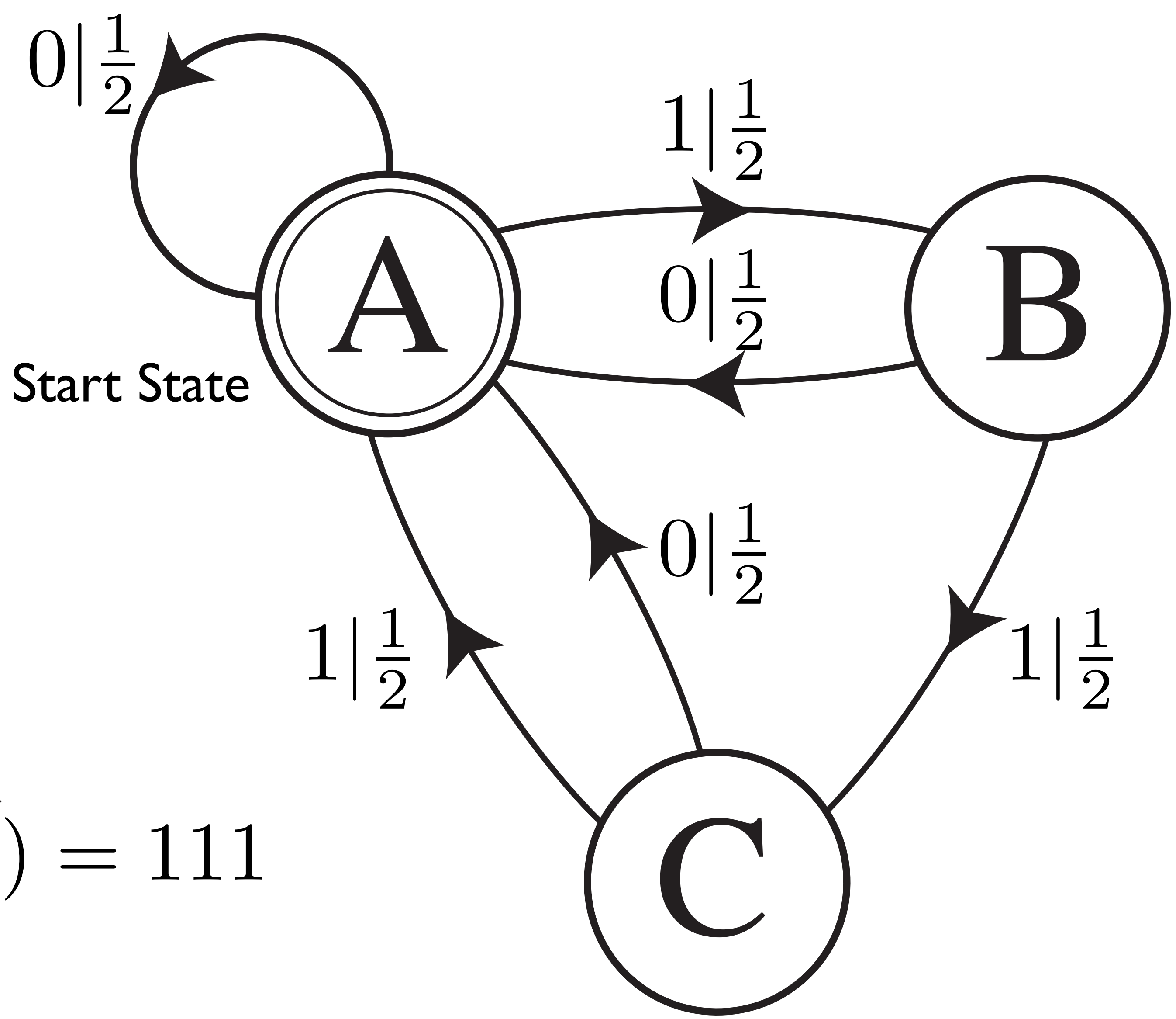
Leaves connect to top tree node

Entropy Rate for Deterministic Hidden Markov Chain ...

Example: Why are modems noisy?

Identify tree nodes with states of a hidden Markov chain





Entropy Rate for Deterministic Hidden Markov Chain ...

Example: Why are modems noisy?

Equivalent hidden Markov chain

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

$$p_V(\infty) = (p_A, p_B, p_C) = (\tfrac{4}{7}, \tfrac{2}{7}, \tfrac{1}{7})$$

It's unifilar:

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

Example: Why are modems noisy?

Calculate entropy rate directly:

$$h_\mu = -\sum_{v \in V} p_v(\infty) \sum_{s \in \mathcal{A}} \sum_{v' \in V} T^{(s)}_{vv'} \log_2 T^{(s)}_{vv'}$$
$$= \tfrac{4}{7} \cdot 1 + \tfrac{2}{7} \cdot 1 + \tfrac{1}{7} \cdot 1$$
$$= 1 \text{ bit}$$

Encoding provides *full* utilization of binary channel.

Modem output sounds noisy!

Entropy Rate for Deterministic Hidden Markov Chain ...

Example: Why are modems noisy?

Compare:

4-symbol source is redundant:

$$\begin{aligned}\mathcal{R} &= \log_2 |\mathcal{A}| - H(X) \\ &= 2 - 1.75 = 0.25 \text{ bits}\end{aligned}$$

Does not use all of 4-symbol channel.

Prefix code mapped 4-symbol, suboptimal source into a new source that uses all available capacity.

Modems do the same: Maximize use of capacity by sending a code stream that is as close to maximum entropy as possible.

Entropy Rate for Nonunifilar Hidden Markov Chain:

Internal: $\{V, T\}$
Observed: $\{T^{(s)} : s \in \mathcal{A}\}$

Entropy rate: No closed-form!

$$h_\mu \neq - \sum_{v \in V} \pi_v \sum_{s \in \mathcal{A}} \sum_{v' \in V} T^{(s)}_{vv'} \log_2 T^{(s)}_{vv'}$$

Upper and Lower Bounds:

$$H(S_L | V_1 S_1 \cdots S_{L-1}) \leq h_\mu(L) \leq H(S_L | S_1 \cdots S_{L-1})$$

Unrealistic for inference: Must know about internal states.
Unrealistic for analysis: Simulate chain, do empirical estimate.

Entropy rate? But there exists a way ... stay tuned!

Information in Processes ...

Motivation:

Previous: Measures of randomness of information source

Block entropy $H(L)$

Entropy rate h_μ

End point of next lectures:

Measures of memory & information storage

Big Picture:

Complementary properties of a source.

Need both: Measures of randomness *and* structure.

How random?

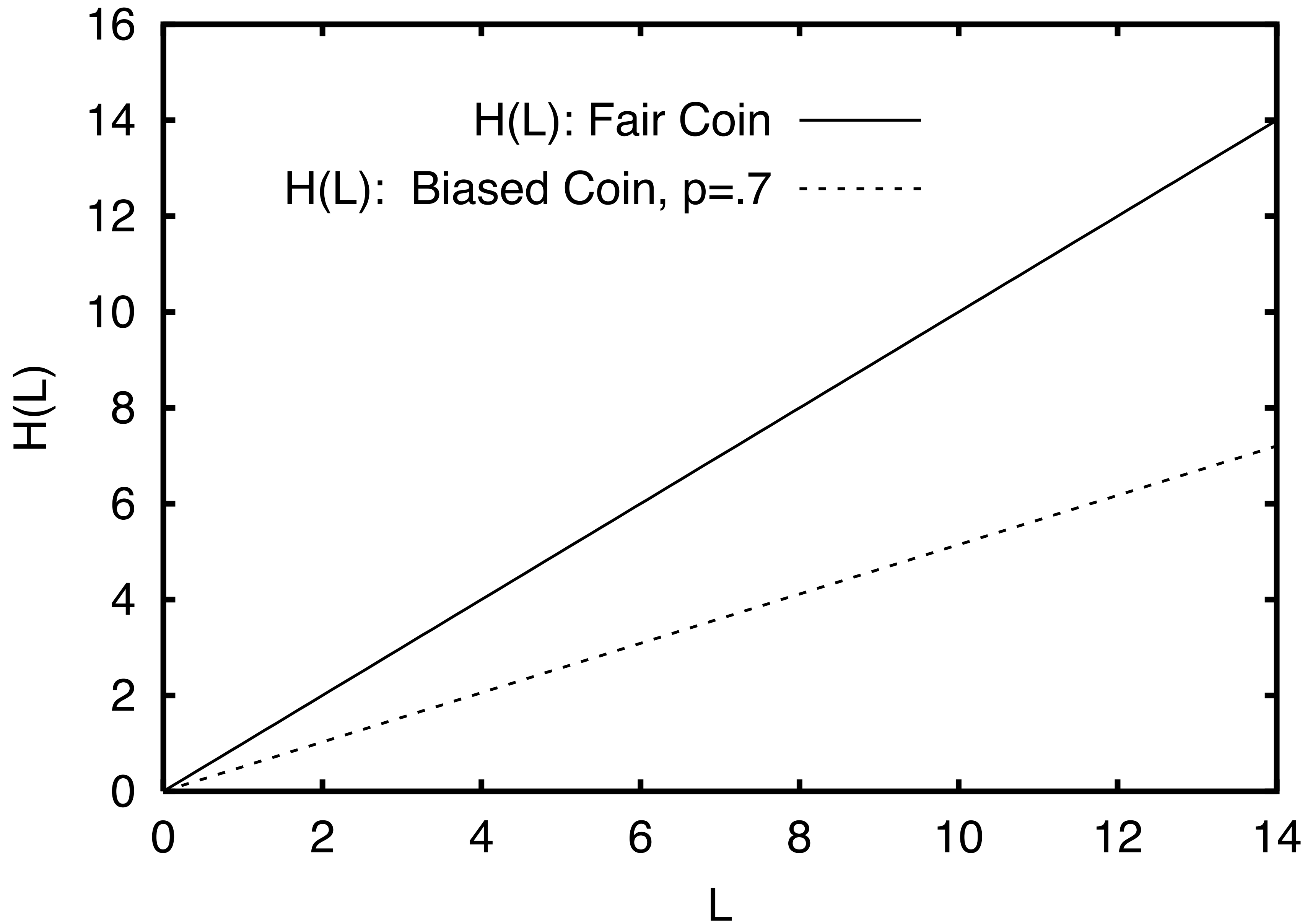
Block entropy growth: $H(L)$.
If L is large enough, we see rate of increase of $H(L)$,
which is the entropy rate:

$$h_\mu = \lim_{L \to \infty} \left(H(L) - H(L-1) \right)$$

Information in Processes ...

How random ...

Fair and Biased Coins:

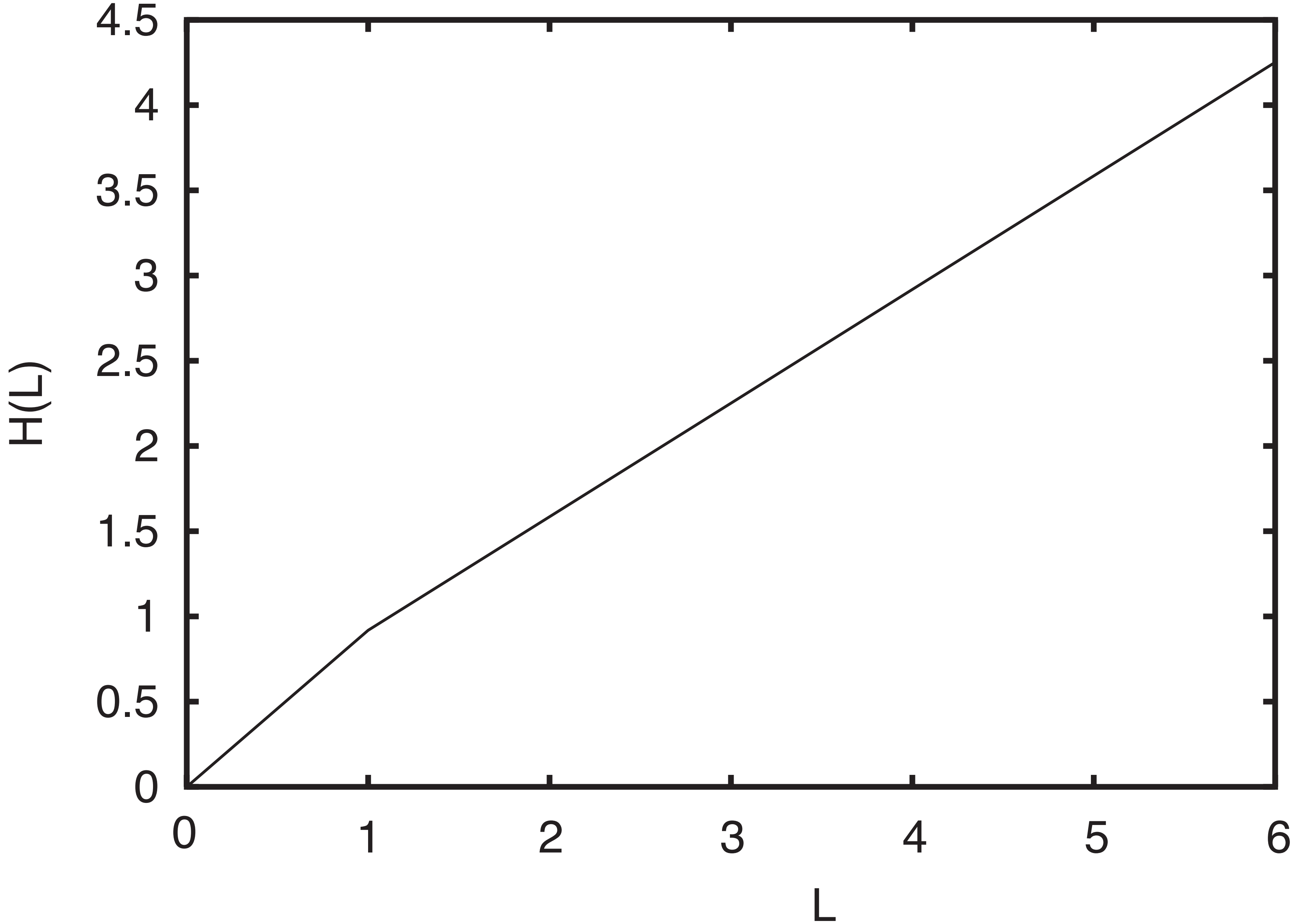


Complexity Lecture 1: Processes and information (CSSS 2011); Jim Crutchfield

Information in Processes ...

How random ...

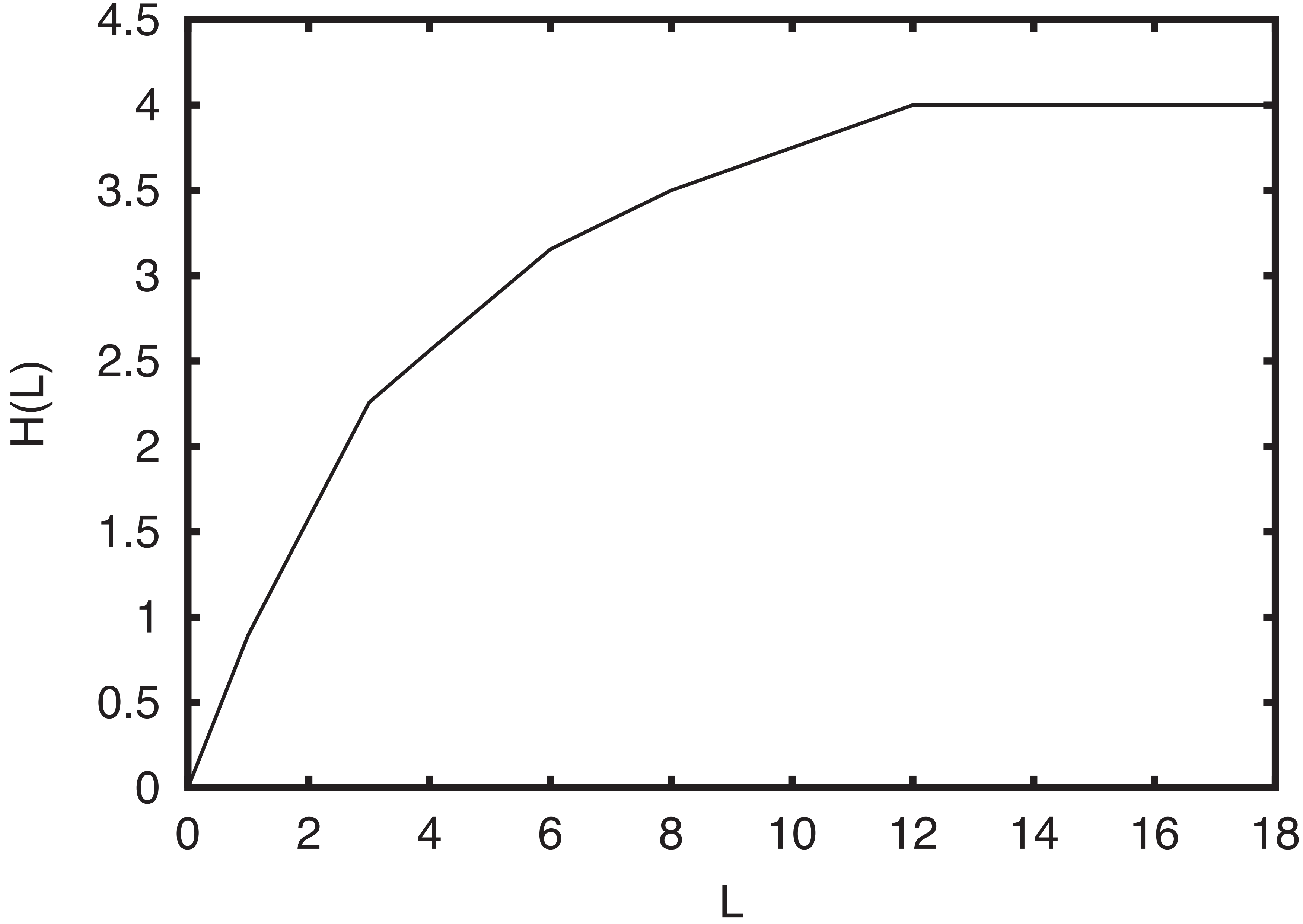
Golden Mean Process:



Information in Processes ...

How random ...

Period-16 Process:

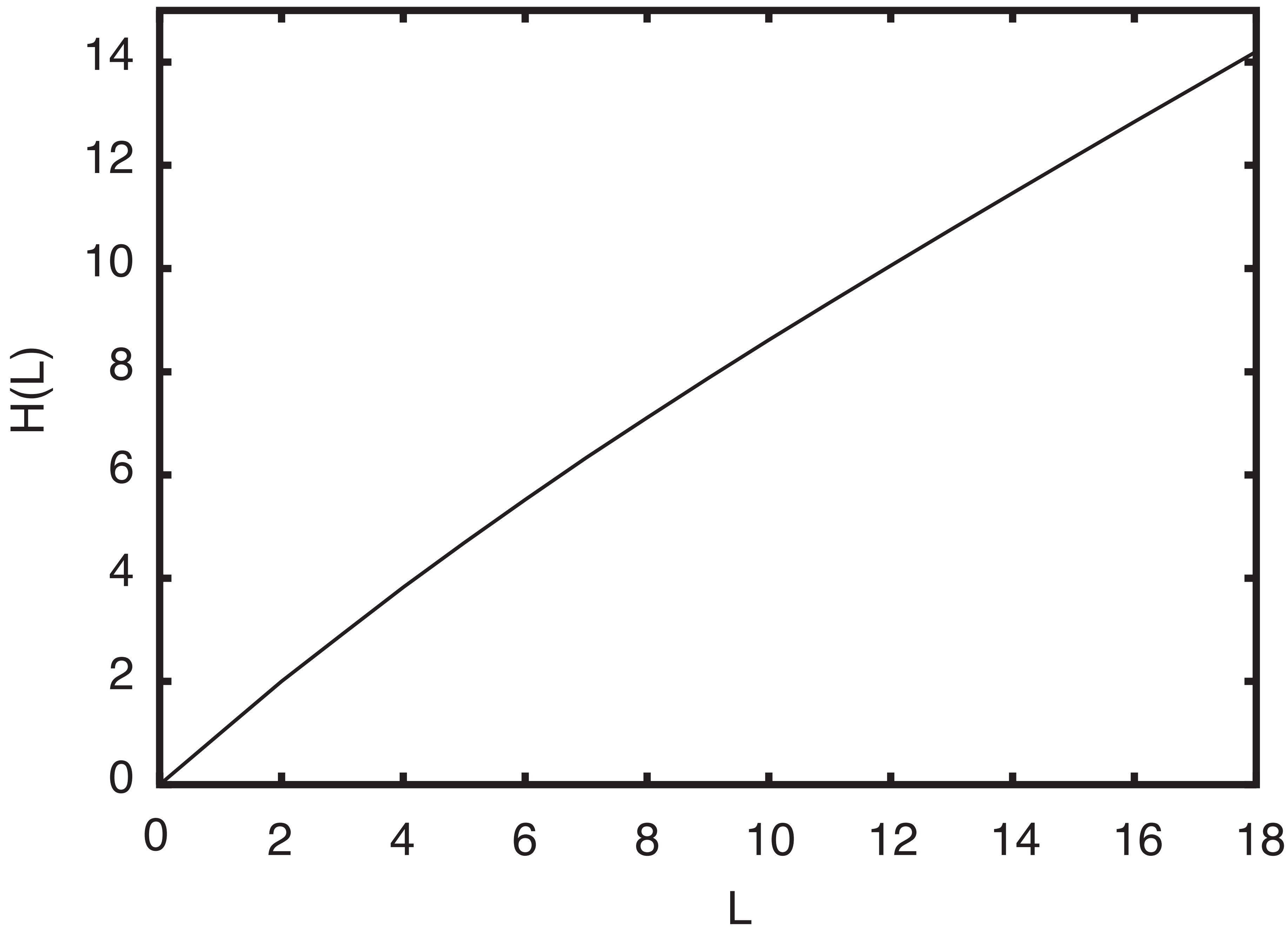


Complexity Lecture 1: Processes and information (CSSS 2011); Jim Crutchfield

Information in Processes ...

How random ...

RRXOR Process:



Complexity Lecture 1: Processes and information (CSSS 2011); Jim Crutchfield

Information in Processes ...

How random ...

How large must L be to see the intrinsic randomness h_μ?

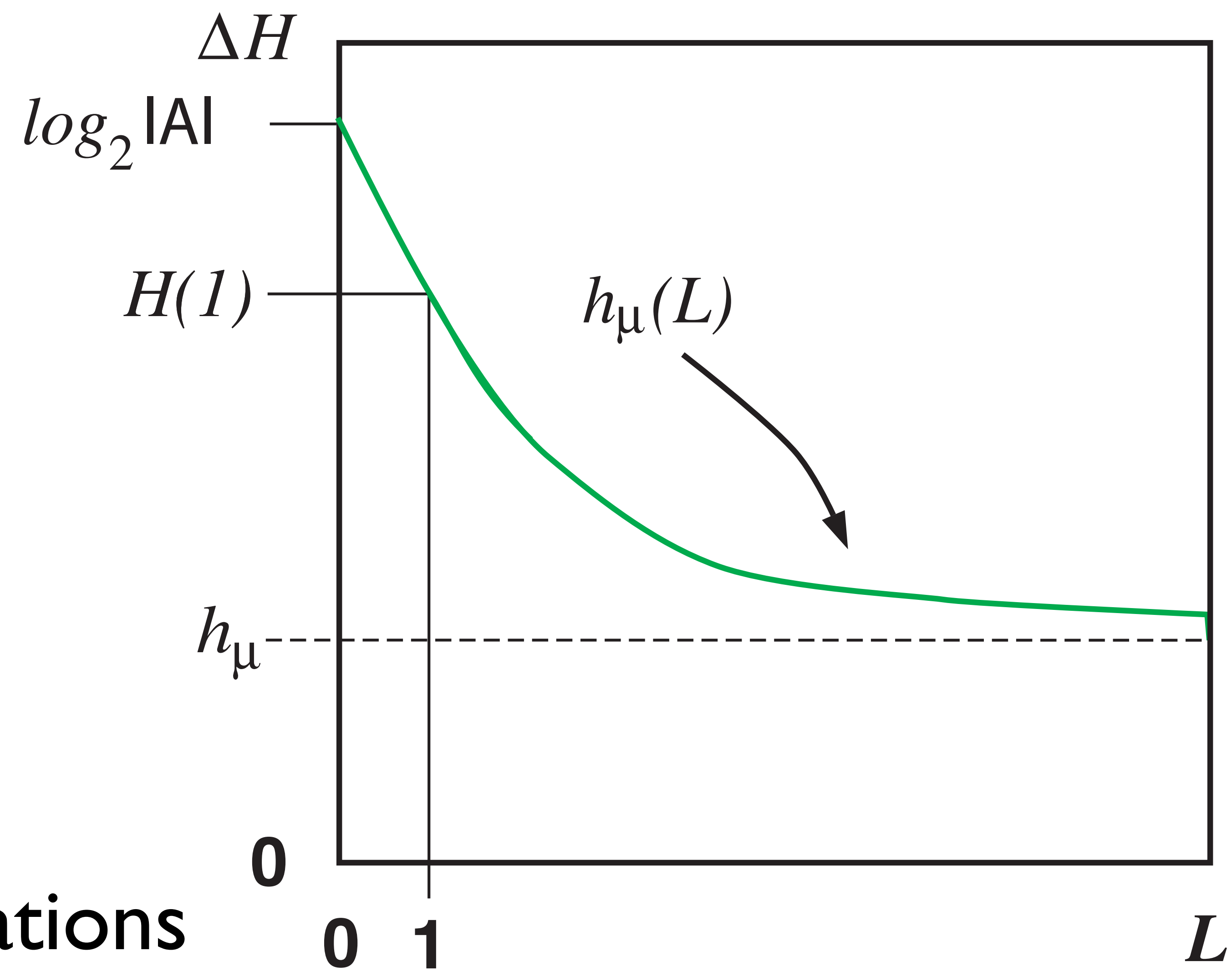
Information in Processes ...

Entropy Convergence:

Length-L entropy rate estimate:

$$h_\mu(L) = \Delta H(L)$$

$$h_\mu(L) = H(L) - H(L-1)$$



Monotonic decreasing:

$$h_\mu(L) \leq h_\mu(L-1)$$

Process appears less random
as account for longer correlations

Entropy (rate) Loss is an Information Gain:

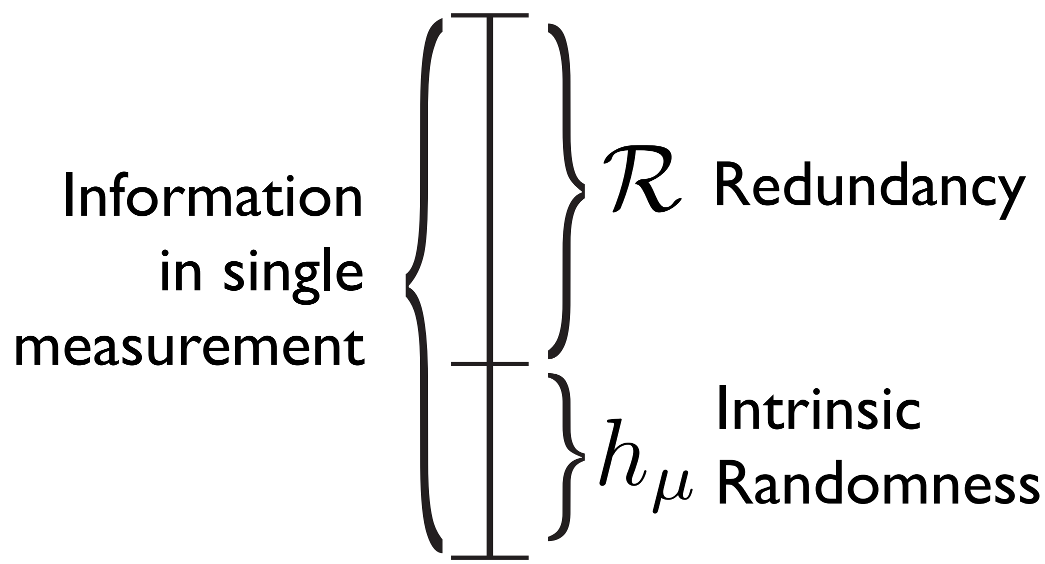
$$h_\mu(L) = \mathcal{D}(\Pr(s^L) || \Pr(s^{L-1}))$$

Information in Processes ...

Redundancy in Processes:

$$\mathcal{R} = \log_2 |\mathcal{A}| - h_\mu$$

Anatomy of Measurement:



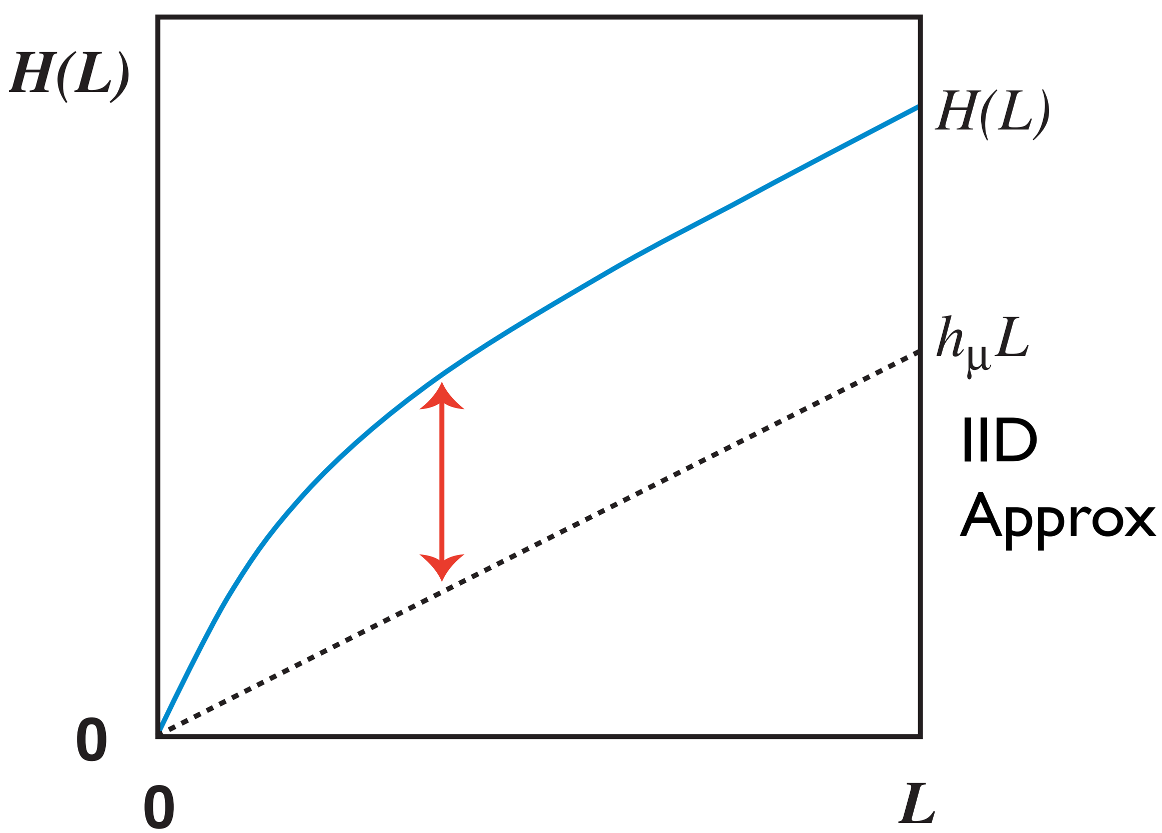
Information in Processes ...

Redundancy in Processes ...

$$\mathcal{R} = \lim_{L \to \infty} \mathcal{D}(\Pr(s^L) || U(s^L))$$

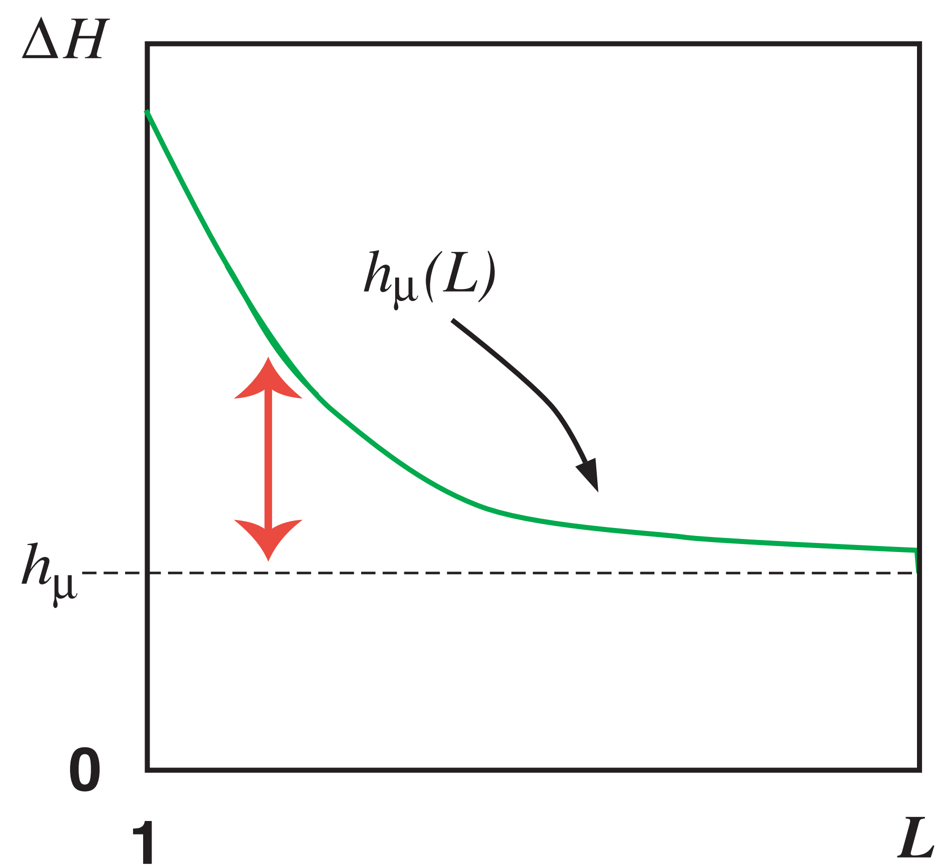
Redundancy in words:

$$\mathcal{R}(L) = H(L) - h_\mu L$$



Redundancy per symbol:

$$r(L) = \mathcal{R}(L) - \mathcal{R}(L-1) = h_\mu(L) - h_\mu$$



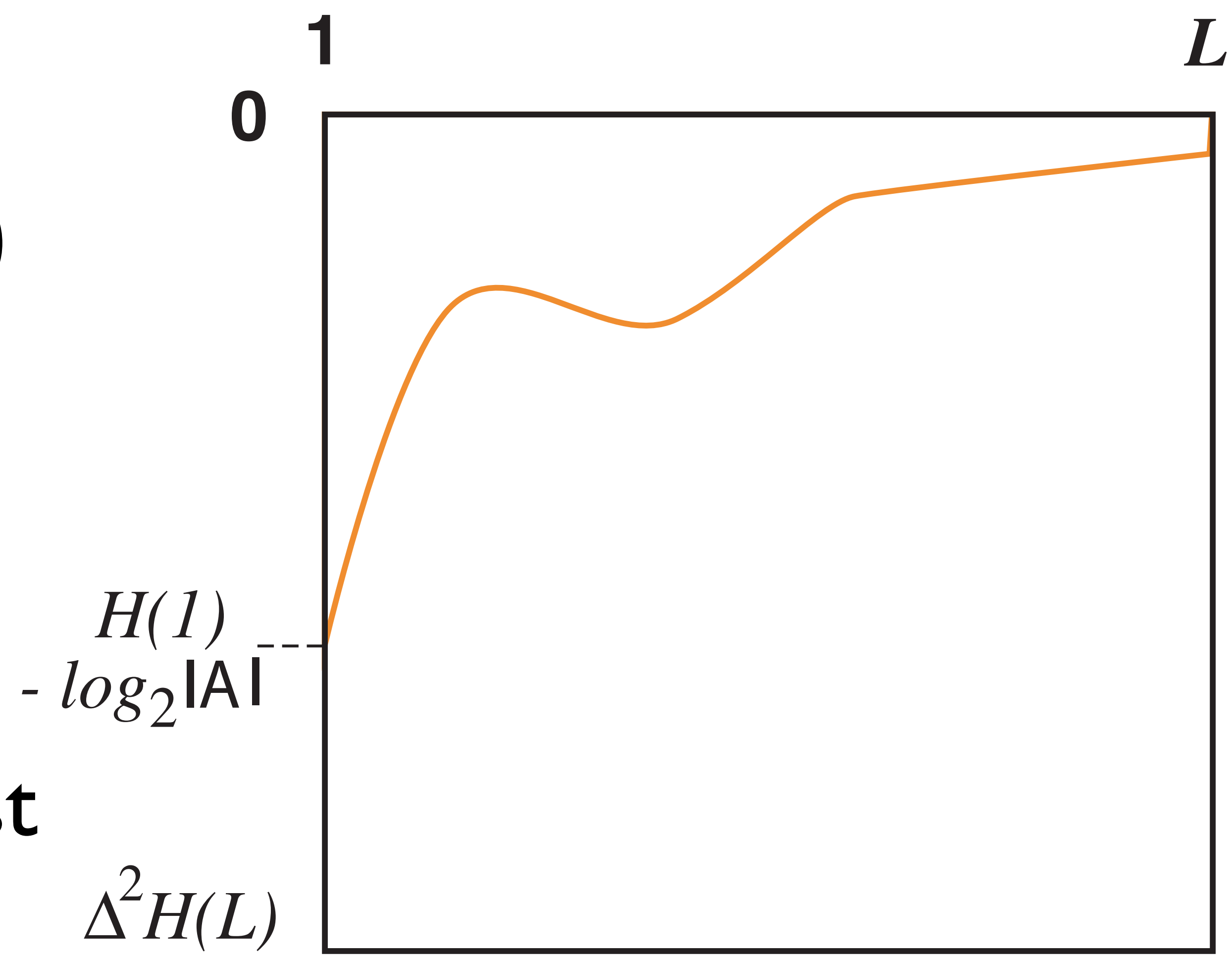
Information in Processes ...

Predictability Gain:

$$\Delta^2 H(L) = h_\mu(L) - h_\mu(L-1)$$

Boundary condition:

$$\Delta^2 H(1) = H(1) - \log_2 |\mathcal{A}|$$



Rate at which unpredictability is lost

Properties:

(1) $H(L)$ Curvature:

$$\Delta^2 H(L) = H(L) - 2H(L-1) - H(L-2)$$

(2) $H(L)$ Concavity:

$$\Delta^2 H(L) \leq 0$$

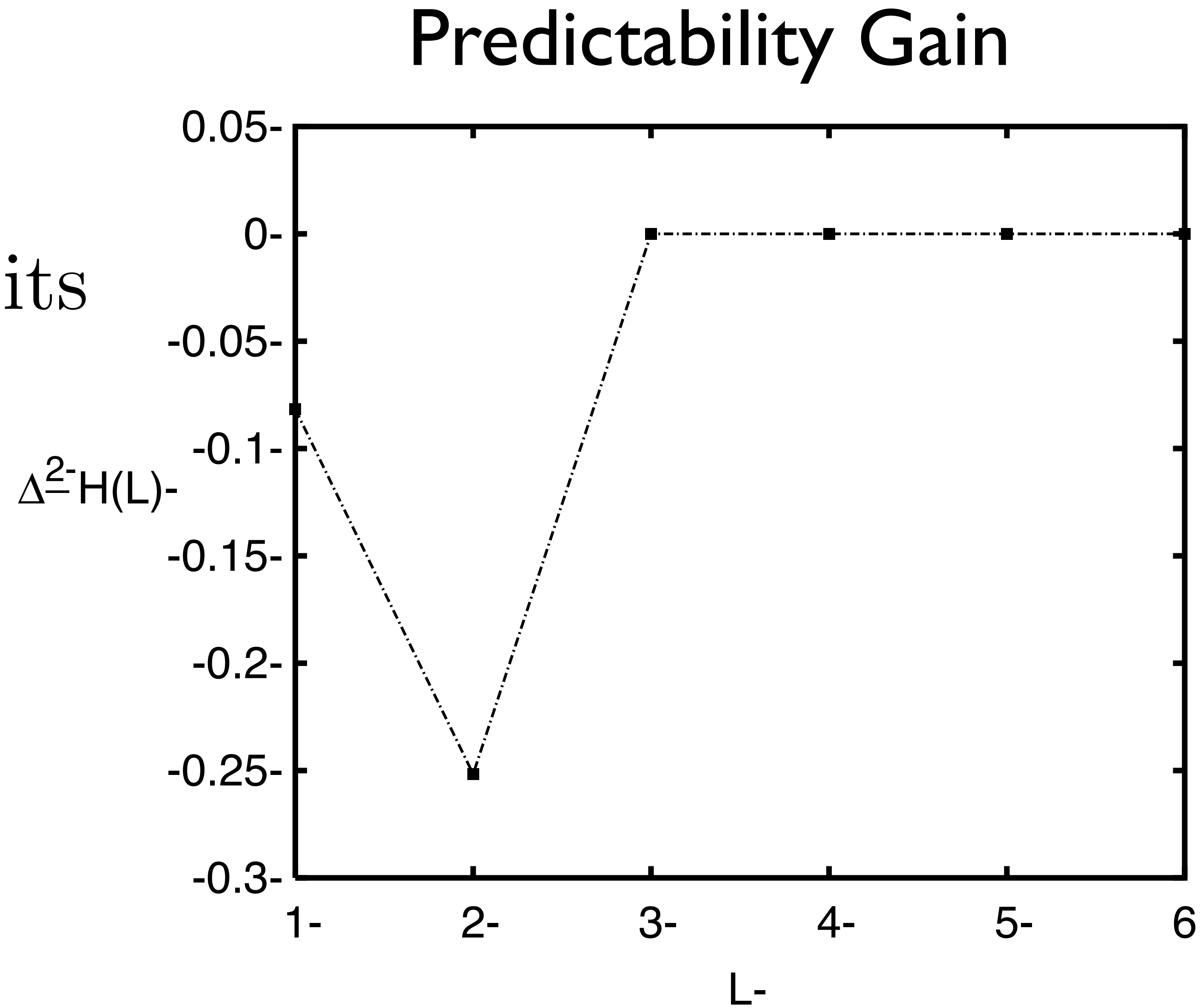
(3) $|\Delta^2 H(L)| \gg 1 \Rightarrow$ Lth measurement significant

Information in Processes ...

Predictability Gain ...

Golden Mean Process:

$$\Delta^2 H(2) = -0.2516 \text{ bits}$$



Second measurement is informative:
00 restriction observed

Entropy Hierarchy:

Take derivatives:

(1) Block entropy: $H(L)$

(2) Entropy rate: $h_\mu(L) = \Delta H(L)$

(3) Predictability gain: $\Delta h_\mu(L) = \Delta^2 H(L)$

Next take integrals!