Complexity

Jim Crutchfield & Ryan James
Complexity Sciences Center
Physics Department
University of California at Davis

Complex Systems Summer School
Santa Fe Institute
St. John’s College, Santa Fe, NM
11 June 2012
The Learning Channel
Last Week

System  Instrument  Process  Modeller

The Learning Channel
The Learning Channel
Morning

System  Instrument  Process  Modeller

The Learning Channel

The Learning Channel
The Learning Channel
The Learning Channel
Complexity

Morning:
  Processes (Jim: 9:30-10:45 AM)
  Information (Jim + Ryan: 11:00 AM-12:00 PM)

Afternoon:
  Structure (Jim: 1:30-2:45 PM)
  Measures of Complexity (Jim: 3:00-4:00 PM)

Evening 6:30-8:00 PM:
  Labs (Ryan)
References? Many, for example:


Applications?
Computational Mechanics

Complexity-Entropy Diagram:

Analyze a class of processes: Chaos, spin systems, biosequences, hydrodynamics, ...

Complex

Structural
Complexity

Simple

Predictable

Randomness
Entropy

Unpredictable

A wide diversity of Complexity-Entropy Diagrams.

Analogous to Thermodynamic Phase Diagram (gas, liquid, solid).
But uses only intrinsic computation properties.


Monday, June 11, 2012
Computational Mechanics: Application to Experimental Molecular Dynamics Spectroscopy

Multiscale complex network of protein conformational fluctuations in single-molecule time series

Chun-Biu Li*, Hao Yang**, and Tamiki Komatsuzaki***

*Nonlinear Sciences Laboratory, Department of Earth and Planetary Sciences, Faculty of Science, Kobe University, Nada, Kobe 657-8501, Japan; **Core Research for Evolutional Science and Technology (CREST), Japan Science and Technology Agency (JST), Kawaguchi, Saitama 332-0012, Japan; ***Department of Chemistry, University of California, Berkeley, CA 94720; and ****Physical Biosciences Division, Lawrence Berkeley National Laboratory, Berkeley, CA 94720


Complexity Lecture 1: Computational Mechanics (CSSS 2012); Jim Crutchfield
Computational Mechanics: Application to Experimental X-Ray Diffraction


Complexity Lecture 1: Computational Mechanics (CSSS 2012); Jim Crutchfield

Monday, June 11, 2012
Measures of Complexity...

Cellular Automata Computational Mechanics

Time

Site

Complexity Lecture 1: Computational Mechanics (CSSS 2012); Jim Crutchfield
Processes and Their Models ...

**Measurement Channel**

Complexity Lecture 1: Processes and information (CSSS 2012) Jim Crutchfield
Main questions now:

How do we characterize the resulting process?

- Measure degrees of unpredictability & randomness.

What correlational structure is there?

How do we build a model from the process itself?

How much can we reconstruct about the hidden internal dynamics?
Stochastic Processes:

Chain of random variables:

\[ \vec{S} \equiv \ldots S_{-2} S_{-1} S_0 S_1 S_2 \ldots \]

Random variable: \( S_t \)

Alphabet: \( \mathcal{A} \)

Realization:

\[ \cdots s_{-2} s_{-1} s_0 s_1 s_2 \cdots ; \ s_t \in \mathcal{A} \]
Processes and Their Models ...

Stochastic Processes:

Chain of random variables: \( S = S_t S_t \)

Past: \( \overleftarrow{S}_t = \ldots S_{t-3} S_{t-2} S_{t-1} \)

Future: \( \overrightarrow{S}_t = S_t S_{t+1} S_{t+2} \ldots \)

L-Block: \( S^L_t \equiv S_t S_{t+1} \ldots S_{t+L-1} \)

Word: \( s^L_t \equiv s_t s_{t+1} \ldots s_{t+L-1} \in \mathcal{A}^L \)
Stochastic Processes ...

Process:
\[ \Pr(S) = \Pr(\ldots S_{-2}S_{-1}S_0S_1S_2\ldots) \]

Sequence (or word) distributions:
\[ \{ \Pr(S_t^L) = \Pr(S_tS_{t+1}\ldots S_{t+L-1}) : S_t \in \mathcal{A} \} \]

Process:
\[ \{ \Pr(S_t^L) : \forall t, L \} \]

Consistency conditions:
\[ \Pr(S_{t+L-1}^L) = \sum_{S_{t+L-1}} \Pr(S_t^L) \quad \Pr(S_{t+1}^{L-1}) = \sum_{S_t} \Pr(S_t^L) \]
Processes and Their Models ...

Types of Stochastic Process:

Stationary process:

$$\Pr(S_t S_{t+1} \ldots S_{t+L-1}) = \Pr(S_0 S_1 \ldots S_{L-1})$$

Assume stationarity, unless otherwise noted.

Notation: Drop time indices.

 Complexity Lecture 1: Processes and information (CSSS 2012) Jim Crutchfield
Models of Stochastic Processes:

**Markov chain model of a Markov process:**

**States:** $v \in A = \{1, \ldots, k\}$

$V = \ldots V_{-2} V_{-1} V_0 V_1 \ldots$

**Transition matrix:** $T_{ij} = \Pr(v_{t+1}|v_t) \equiv p_{vv'}$

$$T = \begin{pmatrix}
p_{11} & \cdots & p_{1k} \\
\vdots & \ddots & \vdots \\
p_{k1} & \cdots & p_{kk}
\end{pmatrix}$$

**Stochastic matrix:** $\sum_{j=1}^{k} T_{ij} = 1$
Models of Stochastic Processes ...

Markov chain ...

Example: \( A = \{ A, B, C \} \)

\[
T = \begin{pmatrix}
p_{AA} & p_{AB} & p_{AC} \\
p_{BA} & p_{BB} & p_{BC} \\
p_{CA} & p_{CB} & p_{CC}
\end{pmatrix}
\]

\[
p_{AA} + p_{AB} + p_{AC} = 1 \\
p_{BA} + p_{BB} + p_{BC} = 1 \\
p_{CA} + p_{CB} + p_{CC} = 1
\]
Processes and Their Models ...

Models of Stochastic Processes ...

Kinds of state:

- **Strongly Connected**
- **Transient**
- **Recurrent**

Complexity Lecture 1: Processes and information (CSSS 2012) Jim Crutchfield
Models of Stochastic Processes ...

Example:
Fair Coin: $\mathcal{A} = \{H, T\}$

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

$$\Pr(H) = \Pr(T) = \frac{1}{2}$$
Processes and Their Models ...

Models of Stochastic Processes ...

Example: Sequence Distribution: \( \Pr(v^L) = 2^{-L} \)

Fair Coin ...

Word as binary fraction:

\[
s^L = s_1 s_2 \ldots s_L
\]

\["s^L" = \sum_{i=1}^{L} \frac{s_i}{2^i}\]

\[s^L \in [0, 1]\]
Processes and Their Models...

Models of Stochastic Processes...

Example:
Fair Coin...

Word as binary fraction:

\[ s^L = s_1 s_2 \ldots s_L \]

\[ \text{"} s^L \text{"} = \sum_{i=1}^{L} \frac{s_i}{2^i} \]

\[ s^L \in [0, 1] \]

Sequence Distribution: \( \Pr(v^L) = 2^{-L} \)
Models of Stochastic Processes ...

Example:

Biased Coin: $\mathcal{A} = \{H, T\}$

\[
T = \begin{pmatrix}
p & 1-p \\
p & 1-p \\
\end{pmatrix}
\]

$\Pr(H) = p$

$\Pr(T) = 1-p$
Examples:

Biased Coin ...

Sequence Distribution:

\[
\Pr(s^L) = p^n(1-p)^{L-n},
\]

\[
n = \text{Number } H \text{s in } s^L
\]
Models of Stochastic Processes ...

Example: Biased Coin ...

Sequence Distribution:

\[ \Pr(s^L) = p^n(1 - p)^{L-n}, \]

\( n = \text{Number } Hs \text{ in } s^L \)
Models of Stochastic Processes ...

Example:

Periodic: \( A = \{A, B, C\} \)

\[
T = \begin{pmatrix}
0 & 1 & 0 \\
0 & 0 & 1 \\
1 & 0 & 0
\end{pmatrix}
\]

Sequence distribution:

\[
Pr(A) = Pr(B) = Pr(C) = \frac{1}{3}
\]

\[
Pr(AB) = Pr(BC) = Pr(CA) = \frac{1}{3} \quad Pr(s^2) = 0 \quad \text{otherwise}
\]

\[
Pr(ABC) = Pr(BCA) = Pr(CAB) = \frac{1}{3} \quad Pr(s^3) = 0 \quad \text{otherwise}
\]
Models of Stochastic Processes ...

Example: Golden Mean Process = “No consecutive 0s”

Markov chain over 2-Blocks: \( \mathcal{A} = \{10, 01, 11\} \)

\[
\begin{pmatrix}
\frac{1}{2} & 0 & \frac{1}{2} \\
\frac{1}{4} & \frac{1}{2} & \frac{1}{4} \\
\frac{1}{4} & \frac{1}{2} & \frac{1}{4}
\end{pmatrix}
\]

\[
\pi = \left( \frac{1}{3}, \frac{1}{3}, \frac{1}{3} \right)
\]

\[
T = \begin{pmatrix}
10 & 01 & 11 \\
\frac{1}{2} & 0 & \frac{1}{2} \\
\frac{1}{4} & \frac{1}{2} & \frac{1}{4} \\
\frac{1}{4} & \frac{1}{2} & \frac{1}{4}
\end{pmatrix}
\]

Complexity Lecture 1: Processes and information (CSSS 2012) Jim Crutchfield
Models of Stochastic Processes …

Example: Golden Mean Process …

Markov chain over 1-Blocks: \( \mathcal{A} = \{0, 1\} \)

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

\[
\pi = \left( \frac{2}{3}, \frac{1}{3} \right)
\]

Also an order-1 Markov chain. Minimal order.

Previous model and this:
Different presentations of the same Golden Mean Process
Processes and Their Models ...

Models of Stochastic Processes ...

Example:

Golden Mean:

\[
\begin{align*}
\text{L} &= 1 \\
\log P &= 5 \\
\text{L} &= 2 \\
\log P &= 5 \\
\text{L} &= 3 \\
\log P &= 5 \\
\text{L} &= 4 \\
\log P &= 5 \\
\text{L} &= 5 \\
\log P &= 5 \\
\text{L} &= 6 \\
\log P &= 5 \\
\text{L} &= 7 \\
\log P &= 5 \\
\text{L} &= 8 \\
\log P &= 5 \\
\text{L} &= 9 \\
\log P &= 5
\end{align*}
\]
Models of Stochastic Processes ...

Two Lessons:

Structure in the behavior: \( \text{supp } \Pr(s^L) \)

Structure in the distribution of behaviors: \( \Pr(s^L) \)
Hidden Markov Models of Processes:

Internal: \( \mathcal{A} = \{A, B, C\} \)

\[
T = \begin{pmatrix}
\rho_{AA} & \rho_{AB} & \rho_{AC} \\
\rho_{BA} & \rho_{BB} & \rho_{BC} \\
\rho_{CA} & \rho_{CB} & \rho_{CC}
\end{pmatrix}
\]

Observed: \( \mathcal{B} = \{0, 1\} \)

\[
T^{(s)} = \begin{pmatrix}
\rho_{AA; s} & \rho_{AB; s} & \rho_{AC; s} \\
\rho_{BA; s} & \rho_{BB; s} & \rho_{BC; s} \\
\rho_{CA; s} & \rho_{CB; s} & \rho_{CC; s}
\end{pmatrix}
\]

\[
\rho_{AA} = \sum_{s \in \mathcal{B}} \rho_{AA; s}
\]

Complexity Lecture 1: Processes and information (CSSS 2012) Jim Crutchfield
Models of Stochastic Processes ...

Hidden Markov Models of Processes ...

Internal states: \( v \in A \)

Transition matrix: \( T = \Pr(v'|v), \ v, v' \in A \)

Observation: Symbol-labeled transition matrices

\[
T^{(s)} = \Pr(v', s|v), \ s \in B
\]

\[
T = \sum_{s \in B} T^{(s)}
\]

Stochastic matrices:

\[
\sum_{j} T_{ij} = \sum_{j} \sum_{s} T_{ij}^{(s)} = 1
\]
Processes and Their Models ...
Models of Stochastic Processes ...
Hidden Markov Models ...

**Internal state distribution:** \( \vec{p}_V = (p_1, p_2, \ldots, p_k) \)

**Evolve internal distribution:** \( \vec{p}_n = \vec{p}_0 T^n \)

**State sequence distribution:** \( v^L = v_0v_1v_2 \ldots v_{L-1} \)

\[
\Pr(v^L) = \pi(v_0)p(v_1|v_0)p(v_2|v_1) \cdots p(v_{L-1}|v_{L-2})
\]

**Observed sequence distribution:** \( s^L = s_0s_1s_2 \ldots s_{L-1} \)

\[
\Pr(s^L) = \sum_{v^L \in A^L} \pi(v_0)p(v_1, s_1|v_0)p(v_2, s_2|v_1) \cdots p(v_{L-1}, s_{L-1}|v_{L-2})
\]

No longer 1-1 map between internal & observed sequences:
Multiple state sequences can produce same observed sequence.
Models of Stochastic Processes ...

Types of Hidden Markov Model:

"Unifilar": current state + symbol “determine” next state

\[
\Pr(v' | v, s) = \begin{cases} 
1 \\ 
0 
\end{cases}
\]

\[
\Pr(v', s | v) = p(s | v)
\]

\[
\Pr(v' | v) = \sum_{s \in A} p(s | v)
\]

“Nonunifilar”: no restriction

Multiple internal edge paths can generate same observed sequence.
Example:

Golden Mean Process as a unifilar HMM:

Internal: \( \mathcal{A} = \{ A, B \} \)

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

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

Observed: \( \mathcal{B} = \{ 0, 1 \} \)

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

Initial ambiguity only: At most 2-to-1 mapping

\( BA^n = 1^n \)  \quad \text{Sync’d: } s = 0 \Rightarrow v = B

\( AA^n = 1^n \)  \quad s = 1 \Rightarrow v = A

Irreducible forbidden words: \( \mathcal{F} = \{ 00 \} \)
Processes and Their Models ...

Models of Stochastic Processes ...

Example:

Golden Mean Process ... Sequence distributions:

Internal state sequences

(A = 1; B = 0)

Observed sequences

Same!

Complexity Lecture 1: Processes and information (CSSS 2012) Jim Crutchfield
Example: Even Process = Even #1s

As a unifilar HMM:

Internal (= GMP): \( \mathcal{A} = \{A, B\} \)

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

\( \mathcal{B} = \{0, 1\} \)

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

\( v^L = \ldots AABAABABA \ldots \)

\( s^L = \ldots 011011110 \ldots \quad s^L = \{\ldots 01^{2n}0 \ldots \} \)

Irreducible forbidden words: \( \mathcal{F} = \{010, 01110, 0111110, \ldots \} \)

No finite-order Markov process can model the Even process!

Lesson: Finite Markov Chains are a subset of HMMs.
Processes and Their Models...

Models of Stochastic Processes...

Example:

Even Process...

Sequence distributions:

Internal states (= GMP)

(A = 1; B = 0)

Observed sequences

Rather different!

Complexity Lecture 1: Processes and information (CSSS 2012) Jim Crutchfield
Processes and Their Models ...

Models of Stochastic Processes ...

Example:

Even Process ...

Sequence distributions:

Internal states (= GMP)  Observed sequences

(A = 1; B = 0)

Rather different!

Complexity Lecture 1: Processes and information (CSSS 2012) Jim Crutchfield
Example:
**Simple Nonunifilar Source:**

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

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

Observed: \( \mathcal{B} = \{0, 1\} \)

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

Many to one: \( 1111111 \Leftrightarrow \{ \text{AAAAAA...} \}
\text{ABBBBBBB...} \text{AABBAAAA...} \text{AAABBBBB...} \ldots \text{BBBBBBBB...} \)

Is there a unifilar HMM presentation of the observed process?
Processes and Their Models ...

Models of Stochastic Processes ...

Example:

Simple Nonunifilar Process ...

Internal states (= Fair coin)  
(A = 1; B = 0)  

Observed sequences
Examples:

Simple Nonunifilar Process ...

Internal states (= Fair coin) $(A = 1; B = 0)$

Observed sequences
Processes and Their Models ...

Models of Stochastic Processes ...

Example:

Simple Nonunifilar Process ...

Internal states (= Fair coin)

\( A = 1; B = 0 \)

Observed sequences

Complexity Lecture 1: Processes and information (CSSS 2012) Jim Crutchfield
What to do with all of this complicatedness?

1. Information theory for general stochastic processes

2. Measures of complexity

3. Optimal models and how to build them

Labs:
Track these topics closely.
Ryan will give a tour in evening session.
Work through them on your own.
Sources of Information:

Apparent randomness:
- Uncontrolled initial conditions
- Actively generated: Deterministic chaos

Hidden regularity:
- Ignorance of forces
- Limited capacity to model structure
Why information?

1. Accounts for any type of co-relation
   • Statistical correlation ~ linear only
   • Information measures nonlinear correlation
2. Broadly applicable:
   • Many systems don’t have “energy”, physical modeling precluded
   • Information defined: social, biological, engineering, ... systems
3. Comparable units across different systems:
   • Distance v. volts v. populations v. energy v. ...
4. Probability theory ~ Statistics ~ Information
5. Complex systems:
   • Emergent patterns!
   • We don’t know these ahead of time
Information as uncertainty and surprise:

Observe something unexpected:
Gain information

Bateson: “A difference that makes a difference”
Information as uncertainty and surprise ...

How to formalize?
Shannon’s approach:
A measure of surprise.
Connection with Boltzmann’s thermodynamic entropy

**Self-information of an event** \( \propto - \log \Pr(\text{event}). \)

Predictable: No surprise \( - \log 1 = 0 \)

Completely unpredictable: Maximally surprised
\[
- \log \frac{1}{\text{Number of Events}} = \log(\text{Number of Events})
\]
Information ...

**Shannon Entropy:** \( X \sim P \)

\[
H(X) = - \sum_{x \in X} p(x) \log_2 p(x)
\]

\[
H(X) = \langle - \log_2 p(x) \rangle
\]

**Units:**
- Log base 2: \( H(X) = [\text{bits}] \)
- Natural log: \( H(X) = [\text{nats}] \)

**Properties:**
- 1. Positivity: \( H(X) \geq 0 \)
- 2. Predictive: \( H(X) = 0 \iff p(x) = 1 \text{ for one and only one } x \)
- 3. Random: \( H(X) = \log_2 k \iff p(x) = U(x) = 1/k \)

Note: \( 0 \log 0 = 0 \)
Examples: Binary random variable $X$

$X = \{0, 1\} \quad \Pr(1) = p \quad \Pr(0) = 1 - p$

$H(X)$?

Binary entropy function:

$$H(p) = -p \log_2 p - (1 - p) \log_2 (1 - p)$$

Fair coin: $p = \frac{1}{2}$

$H(p) = 1 \text{ bit}$

Completely biased coin: $p = 0$ (or 1)

$H(p) = 0 \text{ bits}$

Recall: $0 \cdot \log 0 = 0$
Example: Independent, Identically Distributed (IID) Process over four events

\[ X = \{a, b, c, d\} \quad \text{Pr}(X) = \left(\frac{1}{2}, \frac{1}{4}, \frac{1}{8}, \frac{1}{8}\right) \]

Entropy: \( H(X) = \frac{7}{4} \) bits

Number of questions to identify the event?
- \( x = a? \) (must always ask at least one question)
- \( x = b? \) (this is necessary only half the time)
- \( x = c? \) (only get this far a quarter of the time)

Average number: \( 1 \cdot 1 + 1 \cdot \frac{1}{2} + 1 \cdot \frac{1}{4} = 1.75 \) questions

Interpretation? Optimal way to ask questions.
Example: IID Process over four events ...

Average number: \[ 1 \cdot 1 + 1 \cdot \frac{1}{2} + 1 \cdot \frac{1}{4} = 1.75 \] questions

\[ \Pr(X) = \left( \frac{1}{2}, \frac{1}{4}, \frac{1}{8}, \frac{1}{8} \right) \]
Example: IID Process over four events ...

Query in a different order:

Average number: \(1 \cdot 1 + 1 \cdot \frac{7}{8} + 1 \cdot \frac{6}{7} \approx 2.7\) questions

\[
\Pr(X) = \left(\frac{1}{2}, \frac{1}{4}, \frac{1}{8}, \frac{1}{8}\right)
\]

\[
\Pr(a, b, c) = \left(\frac{4}{7}, \frac{2}{7}, \frac{1}{7}\right)
\]

\[
\Pr(a, b) = \left(\frac{2}{3}, \frac{1}{3}\right)
\]
Example: IID Process over four events

Entropy: $H(X) = \frac{7}{4}$ bits

At each stage, ask questions that are most informative.

Choose partitions of event space that give “most random” measurements.

Theorem:
Entropy gives the smallest number of questions to identify an event, on average.
Interpretations of Shannon Entropy:

Observer’s *degree of surprise* in outcome of a random variable

Uncertainty *in* random variable

Information required to *describe* random variable

A measure of *flatness* of a distribution
Two random variables: \((X, Y) \sim p(x, y)\)

**Joint Entropy**: Average uncertainty in \(X\) and \(Y\) occurring

\[
H(X, Y) = - \sum_{x \in X} \sum_{y \in Y} p(x, y) \log_2 p(x, y)
\]

Independent:

\(X \perp Y \Rightarrow H(X, Y) = H(X) + H(Y)\)
Conditional Entropy: Average uncertainty in X, knowing Y

\[ H(X|Y) = - \sum_{x \in X} \sum_{y \in Y} p(x, y) \log_2 p(x|y) \]

\[ H(X|Y) = H(X, Y) - H(Y) \]

Not symmetric: \( H(X|Y) \neq H(Y|X) \)
Common Information Between Two Random Variables:

\[ X \sim p(x) \& Y \sim p(y) \]

\[ (X, Y) \sim p(x, y) \]

**Mutual Information:**

\[ I(X; Y) = \mathcal{D}(P(x, y)\|P(x)P(y)) \]

\[ I(X; Y) = \sum_{(x, y) \in \mathcal{X} \times \mathcal{Y}} p(x, y) \log_2 \frac{p(x, y)}{p(x)p(y)} \]
Information ...

Mutual Information ...

Properties:

1. \( I(X; Y) \geq 0 \)
2. \( I(X; Y) = I(Y; X) \)
3. \( I(X; Y) = H(X) - H(X|Y) \)
4. \( I(X; Y) = H(X) + H(Y) - H(X, Y) \)
5. \( I(X; X) = H(X) \)
6. \( X \perp Y \Rightarrow I(X; Y) = 0 \)

Interpretations:

- Information one variable has about another
- Information shared between two variables
- Measure of dependence between two variables
Event Space Relationships of Information Quantifiers:
Event Space Relationships of Information Quantifiers:
Event Space Relationships of Information Quantifiers:

\[ H(X) \]
Event Space Relationships of Information Quantifiers:

\[ H(X) \]
Event Space Relationships of Information Quantifiers:

\[ H(Y) \quad H(X) \]
Information ...

Event Space Relationships of Information Quantifiers:

\[ H(X, Y) \]

\[ H(Y) \]

\[ H(X) \]
Event Space Relationships of Information Quantifiers:

\[ H(X, Y) \]

\[ H(Y) \]

\[ H(X|Y) \]

\[ H(X) \]
Event Space Relationships of Information Quantifiers:

\[ H(X, Y) \]

\[ H(Y) \]

\[ H(Y|X) \]

\[ H(X|Y) \]

\[ H(X) \]
Information ...

Event Space Relationships of Information Quantifiers:

\[ H(X, Y) \]

\[ H(Y) \] \hspace{1cm} \[ H(Y|X) \] \hspace{1cm} \[ I(X;Y) \] \hspace{1cm} \[ H(X|Y) \] \hspace{1cm} \[ H(X) \]
Event Space Relationships of Information Quantifiers:

- $H(X,Y)$
- $H(Y)$
- $H(Y|X)$
- $I(X;Y)$
- $H(X|Y)$
- $H(X)$
- $d(X;Y)$

Complexity Lecture 1: Processes and Information theory (CSSS 2012); Jim Crutchfield
Three random variables: \((X, Y, Z) \sim p(x, y, z)\)

Markov Chain: \(X \rightarrow Y \rightarrow Z\)

\[
p(x, z|y) = p(x|y)p(z|y)
\]

or

\[
I(X; Z|Y) = 0
\]

Y shields \(X\) and \(Z\) from each other: \(X \perp_Y Z\)

Properties:

(1) \(X \rightarrow Y \rightarrow Z \Rightarrow Z \rightarrow Y \rightarrow X\)

(2) \(Z = f(Y) \Rightarrow X \rightarrow Y \rightarrow Z\)
Data Processing Inequality:

\[ X \rightarrow Y \rightarrow Z \Rightarrow I(X; Y) \geq I(X; Z) \]

Corollary:

\[ Z = g(Y) \Rightarrow I(X; Y) \geq I(X; g(Y)) \]

Manipulation cannot increase information about \( X \).
Real Information Theory:

How to compress a process:
Can’t do better than $H(X)$
(Shannon’s First Theorem)

How to communicate a process’s data:
Can transmit error-free at rates up to channel capacity
(Shannon’s Second Theorem)

Both results give operational meaning to entropy.
Previously, entropy motivated as a measure of surprise.
Information in Processes ...

Communication channel:

\[ \ldots x_3x_2x_1 \quad \ldots C(x_3)C(x_2)C(x_1) \quad \ldots \hat{C}(x_3)\hat{C}(x_2)\hat{C}(x_1) \quad \ldots x_3x_2x_1 \]

- **Messages**
- **Codewords**
- **Corrupted Codewords**
- **Inferred Messages**

Complexity Lecture 1: Processes and Information theory (CSSS 2012); Jim Crutchfield
Codebook: $C$

**Code rate:** $R(C) = \text{number bits per message.}$

**Data Compression Theorem (Shannon’s First Theorem):**

$$R(C) \geq H(X)$$

Cannot compress source below its entropy rate.

**Operational meaning of entropy:** fundamental limit.
Information in Processes ...
Coding for Communication Channels ...

**Discrete channel:**

- **Input:** \( X \sim p(x) \)
- **Output:** \( Y \sim p(y) \)
- **Channel:** \( p(y|x) \)

**Memoryless channel:**

\[
p(y_t|x_tx_{t-1} \cdots) = p(y_t|x_t)
\]

**Channel Capacity:**

\[
C = \max_{p(x)} I(X;Y)
\]
Channel Coding Theorem (Shannon’s Second Theorem):

1. Capacity is the maximum reliable transmission rate.
2. Error-free codes exist if $R < C$.

Idea:
Model as noisy channel with non-overlapping outputs.

Strategy:
Code long block lengths: $|\mathcal{X}^L| \approx 2^{LH(X)}$
Choose codewords (channel inputs) that produce non-overlapping outputs.
What happens when transmitting above capacity, $R > C$?

(Typical of measurement systems?)