How Random is a Coin Toss?

Bayesian Inference and the Symbolic Dynamics of Deterministic Chaos

Christopher C. Strelioff [1,2]   Dr. James P. Crutchfield [2]

[1] Center for Complex Systems Research and Department of Physics
University of Illinois at Urbana-Champaign

[2] Computational Science and Engineering Center and Department of Physics
University of California at Davis

NIPS 2006

Workshop on Dynamical Systems, Stochastic Processes and Bayesian Inference

December 9th, 2006

Sources of Randomness

Stochastic or Deterministic?

- Given a data source which appears to be a high-dimensional stochastic process:
  - Is this instead a low-dimensional deterministic system which exhibits a chaotic attractor?
- Analysis of chaotic data:
  - Chaotic data is very sensitive to the *measurement* process.
  - This is both a blessing and a curse.

A model of measurement

Symbolic Dynamics

- Finite-precision instrument:
  - Induces a symbolic representation.
  - How well does this represent underlying continuous-valued behavior?
- *Symbolic dynamics:*
  - Coarse-grained view of continuous dynamics.
  - Use ideas from symbolic dynamics to discuss *instrument design*.

Modeling chaotic time series

Continuous-state → discrete data

- Instrument design - *Maximize entropy rate*:
  - Project continuous state onto finite set of disjoint regions.
  - Produce maximum information with each measurement.
- Model inference - *Minimize entropy rate*:
  - Model discrete data produced by instrument.
  - Search for determinism and structure in data.

Our model data source

1-dimensional chaotic map

- 1-d map with additive noise:

$$x_{t+1} = f(x_t) + \xi_t$$

where

$$x_t \in [0, 1], \quad \xi_t \sim \mathrm{N}(0, \sigma^2).$$

- Why is this a relevant example?
  - We can connect ordinary differential equations with a map.

Rössler Attractor

How can we treat this ODE as a 1-d map?

- The system of ode's

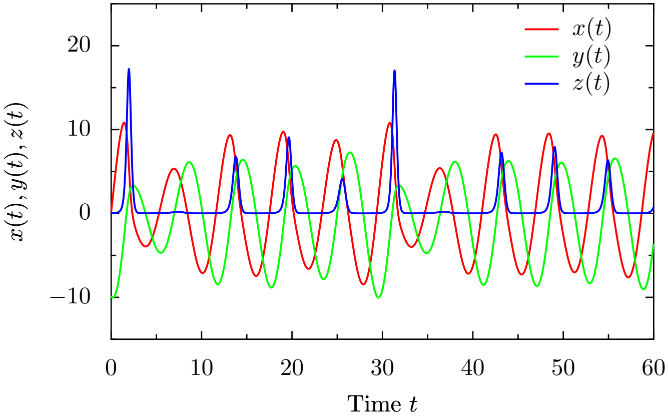
$$\begin{aligned} \frac{dx}{dt} &= -y - z \\ \frac{dy}{dt} &= x + ay \\ \frac{dz}{dt} &= b + z(x - c) \end{aligned}$$

- Typical chaotic parameters:

$$a = 0.2, b = 0.2, c = 5.7$$

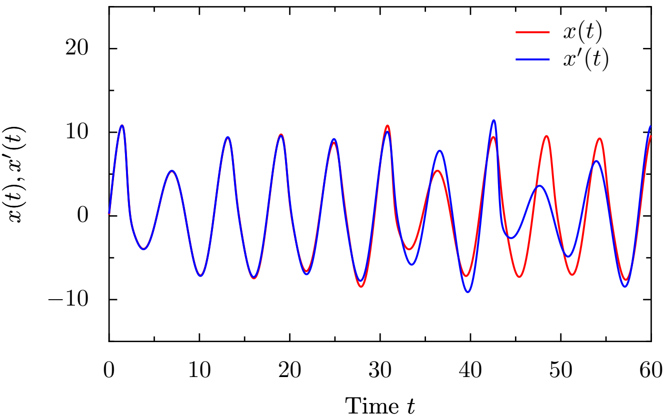
Rössler Attractor

Representative chaotic dynamics



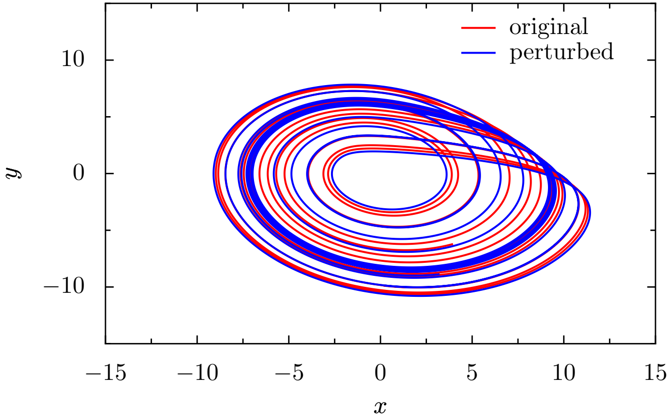
Rössler Attractor

Randomness - sensitivity to perturbation



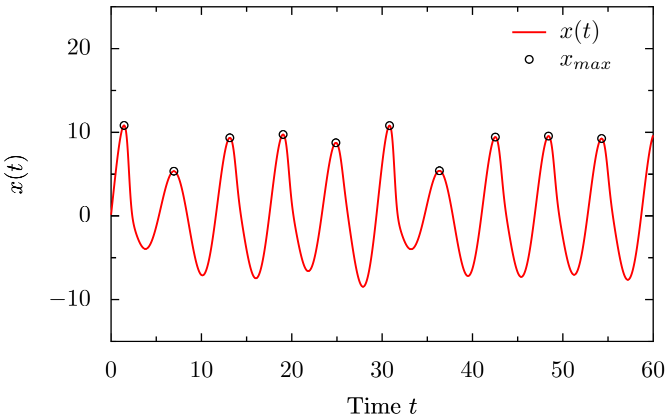
Rössler Attractor

Typical Chaotic Dynamics



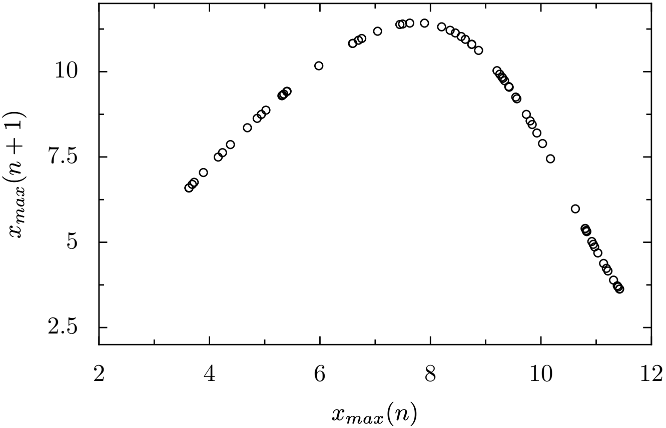
Rössler Attractor

Consider the sequence of x_{max}



Rössler Attractor

1-d map from x_{max} sequence



Designing an instrument

Symbolic dynamics: without noise

- Assume a discrete time map (state space M):

$$f : M \to M$$

- Partition M into a finite number of non-overlapping regions

$$\mathcal{P} = \{I_i : \cup_i I_i = M, I_i \cap I_j = \emptyset, i \neq j\}$$

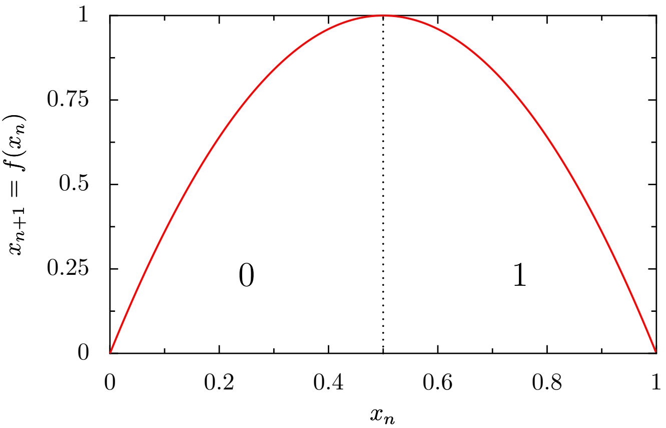
- For one-dimensional maps:
    - Choose *decision points* at *critical points* of map.

Chris Strelioff
James Crutchfield

Symbolic Dynamics

Logistic map: $x_{n+1} = r\, x_n\, (1 - x_n)$

- A good partition:



Choose a partition $\mathcal{P}$

Create coarse-grained data

- What if we don't know $f(x)$?
- Choose a $\mathcal{P}$ and create symbolic sequence:

$$\mathbf{X} = x_0 x_1 \ldots x_{N-1} \underset{\mathcal{P}}{\rightarrow} \mathbf{S} = s_0 s_1 \ldots s_{N-1} \quad , \; s_t \in \mathcal{A}.$$

- Best partition gives true entropy rate

$$h_\mu = \max_{\{\mathcal{P}\}} h_\mu(\mathcal{P}) \; .$$

Testing the resulting partition

Lyapunov Exponent

- *Piesin's Identity*

$$h_\mu = \sum_i \lambda_i^+$$

- *Lyapunov Exponent*

$$\lambda = \lim_{N \to \infty} \frac{1}{N} \sum_{t=1}^{N} \log_2 |f'(x_t)| \ .$$

Bayes' Theorem

Symbolic data → Markov model

- Goal is to find posterior:

$$P(\theta_k|D, M_k) = \frac{P(D|\theta_k, M_k)P(\theta_k|M_k)}{P(D|M_k)}$$

- M_k: Markov chain order-k
- θ_k: model parameters
- D: data

Likelihood

k-th order Markov Chains

- Likelihood

$$P(D|\theta_k, M_k) = \prod_{s \in \mathcal{A}} \prod_{\overleftarrow{s}^k \in \mathcal{A}^k} p(s|\overleftarrow{s}^k)^{n(\overleftarrow{s}^k s)}$$

  - $n(\overleftarrow{s}^k s)$: counts of word $\overleftarrow{s}^k s$ in data D.

- Assumes:
  - Finite memory.
  - Stationarity.

Prior

k-th order Markov Chains

- Prior is product of Dirichlet distributions

$$P(\theta_k|M_k) = \prod_{\overleftarrow{s}^k \in \mathcal{A}^k} \left\{ \frac{\Gamma(\alpha(\overleftarrow{s}^k))}{\prod_{s\in\mathcal{A}} \Gamma(\alpha(\overleftarrow{s}^k s))} \times \delta(1 - \sum_{s\in\mathcal{A}} p(s|\overleftarrow{s}^k)) \prod_{s\in\mathcal{A}} p(s|\overleftarrow{s}^k)^{\alpha(\overleftarrow{s}^k s)-1} \right\}$$

- where

$$\alpha(\overleftarrow{s}^k) = \sum_s \alpha(\overleftarrow{s}^k s)$$

Evidence

k-th order Markov Chains

- Evidence or marginal likelihood

$$
\begin{aligned}
P(D|M_k) &= \int d\theta_k \, P(D|\theta_k, M_k) P(\theta_k|M_k) \\
&= \prod_{\overleftarrow{s}^k \in \mathcal{A}^k} \left\{ \frac{\Gamma(\alpha(\overleftarrow{s}^k))}{\prod_{s \in \mathcal{A}} \Gamma(\alpha(\overleftarrow{s}^k s))} \right. \\
&\times \left. \frac{\prod_{s \in \mathcal{A}} \Gamma(n(\overleftarrow{s}^k s) + \alpha(\overleftarrow{s}^k s))}{\Gamma(n(\overleftarrow{s}^k) + \alpha(\overleftarrow{s}^k))} \right\}
\end{aligned}
$$

- This term is fundamental to model comparison, estimation of h_μ.

Posterior

k-th order Markov Chains

- Combine likelihood, prior, evidence using Bayes' theorem to obtain posterior density

$$P(\theta_k|D,M_k) = \prod_{\overleftarrow{s}^k \in \mathcal{A}^k} \left\{ \frac{\Gamma(n(\overleftarrow{s}^k) + \alpha(\overleftarrow{s}^k))}{\prod_{s\in\mathcal{A}} \Gamma(n(\overleftarrow{s}^k s) + \alpha(\overleftarrow{s}^k s))} \right.$$

$$\left. \times \delta(1 - \sum_{s\in\mathcal{A}} p(s|\overleftarrow{s}^k)) \prod_{s\in\mathcal{A}} p(s|\overleftarrow{s}^k)^{n(\overleftarrow{s}^k s)+\alpha(\overleftarrow{s}^k s)-1} \right\}$$

Model comparison

How do we select the order k?

- Consider set of orders $\mathcal{M} = \{M_k\}$
- Probability of order k is given by

$$P(M_k|D,\mathcal{M}) = \frac{P(D|M_k,\mathcal{M})\,P(M_k|\mathcal{M})}{\sum_{M_{k'}} P(D|M_{k'},\mathcal{M})\,P(M_{k'}|\mathcal{M})}$$

- Penalize for model size

$$P(M_k|\mathcal{M}) = \exp(-|\mathcal{A}|^k(|\mathcal{A}|-1))$$

Estimating h_μ

Define a partition function

- Evidence becomes partition function

$$\mathcal{Z} = \int d\theta_k \; P(D|\theta_k, M_k) P(\theta_k|M_k)$$

- Re-write product of likelihood and prior

$$P(D, \theta_k|M_k) \propto 2^{-\beta_k(\mathcal{D}[Q\|P]+h_\mu[Q])} 2^{+|\mathcal{A}|^{k+1}(\mathcal{D}[U\|P]+h_\mu[U])}$$

  - where $\beta_k = \sum_{\overleftarrow{s}^k, s} n(\overleftarrow{s}^k s) + \alpha(\overleftarrow{s}^k s)$.

- Distributions

  - Q: posterior mean, U: uniform.

Estimating h_μ

Define a partition function

- An average of $\mathcal{D}[Q\|P] + h_\mu[Q]$

$$-\frac{1}{\log 2}\frac{\partial}{\partial \beta_k} log \mathcal{Z} = \mathbf{E}_{post}\left[\,\mathcal{D}[Q\|P] + h_\mu[Q]\,\right]$$

- Analytic result

$$\mathbf{E}_{post}\left[\,\mathcal{D}[Q\|P] + h_\mu[Q]\,\right] = \frac{1}{\log 2}\sum_{\overleftarrow{s}^k} q(\overleftarrow{s}^k)\psi^{(0)}\left[\beta_k q(\overleftarrow{s}^k)\right]$$

$$-\frac{1}{\log 2}\sum_{\overleftarrow{s}^k, s} q(\overleftarrow{s}^k) q(s|\overleftarrow{s}^k)\psi^{(0)}\left[\beta_k q(\overleftarrow{s}^k) q(s|\overleftarrow{s}^k)\right]$$

- where $\psi^{(0)}$ is the digamma function.

Chris Strelioff
James Crutchfield

An experiment

Instrument and Inference

- We know how to create an instrument - define $\mathcal{P}$.
  - *GOAL*: Maximize h_μ - informative measurements.
- We know how to infer a Markov chain model of coarse grained data.
  - *GOAL*: Minimize h_μ - find patterns.

Methods

Our experiment

- Generate a *single* time series of length $N = 10^4$

$$f(x_t) = r\, x_t\, (1 - x_t) \qquad , r = 4.0,\ \sigma = 10^{-3}$$

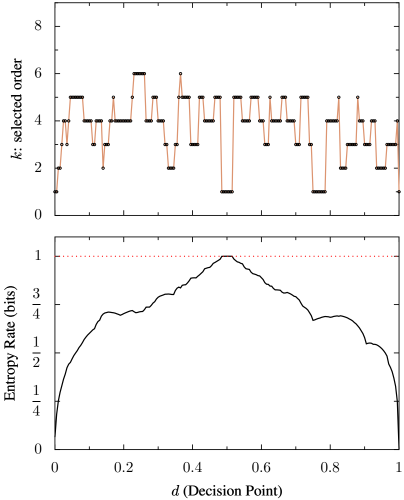
- Design an instrument
  - Choose a decision point d

$$\mathcal{P}(d) = \{\text{“0”} \sim x \in [0, d), \text{“1”} \sim x \in [d, 1]\}$$

- Infer a Markov chain and estimate h_μ.
  - Consider orders $k = 1 - 8$.

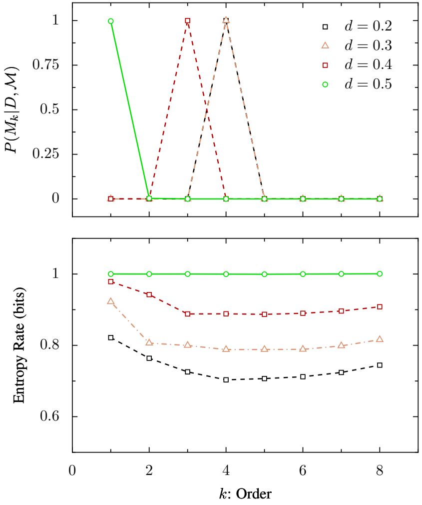
Instrument design

Maximize h_μ



Model Inference

Minimize h_μ



Conclusion

Bayesian Inference and the Symbolic Dynamics of Deterministic Chaos

- Randomness can be due to low-dimensional chaos.
- Symbolic dynamics can help analyze this data.
- For instruments, maximize observed entropy rate.
- For model inference, minimize entropy rate.