Discovering Hidden Markov Models from Time Series

Cosma Shalizi
(Complex Systems, University of Michigan)

Hidden State Models for Time Series

- What we can see is bad (non-Markovian, non-stationary, etc.)
- Hidden state: what we *can't* see is nice
- Usually: guess structure, see if it works
  - EM algorithm for parameters + states
  - Bayesian updating for state estimation
- State-space reconstruction
  - Entirely data-driven
  - No EM or Bayes needed
  - No good with stochastic dynamics

Markovian Representations

(Knight, 1975)

- Any process is a noisy function of a homogeneous Markov process

  Markovian states = conditional distributions

- Can we find the Markovian states?

Causal States

(Crutchfield & Young, 1989)

- past A and past B equivalent iff

$$\Pr(\text{Future}|A) = \Pr(\text{Future}|B)$$

- [A] = all pasts equivalent to A
- Statistic ("causal state"):

$$\epsilon(\text{past}_t) = [\text{past}_t] = s_t$$

- Each state ≡ conditional distribution
- IID = 1 state, periodic = p states

Markov Properties

(Shalizi & Crutchfield, 2001)

$$\text{future}_t \perp \text{past}_t \mid s_t$$

- Recursive transitions for states

$$\epsilon(\text{past}_t\, x_{t+1}) = f(\epsilon(\text{past}_t), x_{t+1})$$

  - No Bayesian updating needed

- $\therefore$ States are Markovian

$$s_{t+1} \perp s_{t-1} \mid s_t$$

Optimality Properties

(Shalizi & Crutchfield, 2001)

- Sufficiency:

  $$I[\text{future};\text{past}] = I[\text{future};\epsilon(\text{past})]$$

  - ∴ Optimal under any loss function

- Minimality: Can compute $\epsilon(\text{past})$ from any other sufficiency statistic
- Uniqueness: no other minimal sufficient statistic
- Minimal stochasticity

HMMs are FSMs

- Assume discrete valued series from now on
- Every finite HMM specifies a finite regular language

States ⇔ States
Observations ⇔ Symbols
Recursive updating ⇔ Deterministic transitions
Positive sequence probability ⇔ Word in language

CSSR

(Causal State Splitting Reconstruction)

with K. L. Shalizi and J. P. Crutchfield

- Key observation:

Recursion + next-step sufficiency

$\Rightarrow$ general sufficiency

- Get next-step distribution right
- Then make states recursive

- Given: a set of states
- See if conditioning on state plus one extra symbol makes a difference
- If yes, sub-divide states
- New conditional distributions?
  - If no, shift state boundaries
  - If yes, make new states
- Start with one state (as if IID)
- Stop when no changes or reach maximum history length

Recursion

- Do all the histories in a state make the same transition on the same symbol?
- If not, split the state
- Keep checking until no state needs to be split

Time Complexity

- One pass through data
- n data points, k symbols, max. length L
- Everything-goes-wrong upper bound

$$O(n) + O(k^{2L+1})$$

Convergence of CSSR

- S = true causal state structure
- S(n) = structure inferred from n data-points
- D = true distribution, D(n) = inferred
- Assume: finite # of states, every state has a finite history, using long enough histories

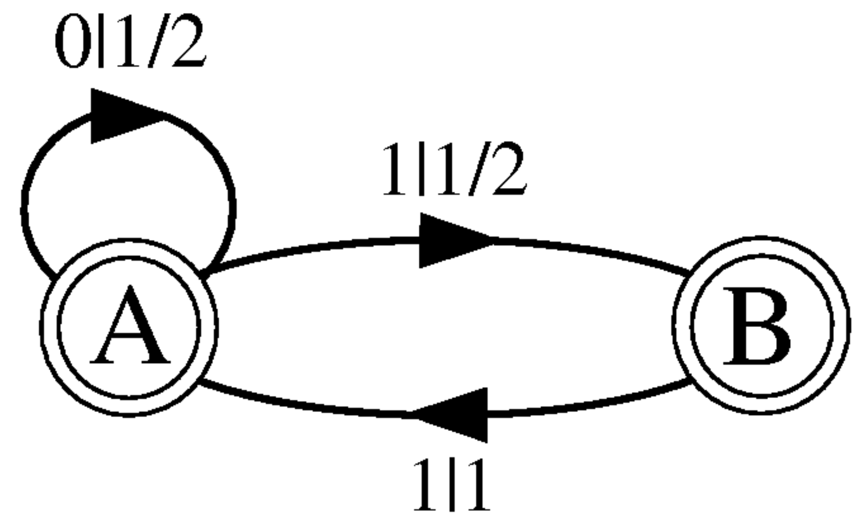
Prob(S(n) ≠ S) Õ 0

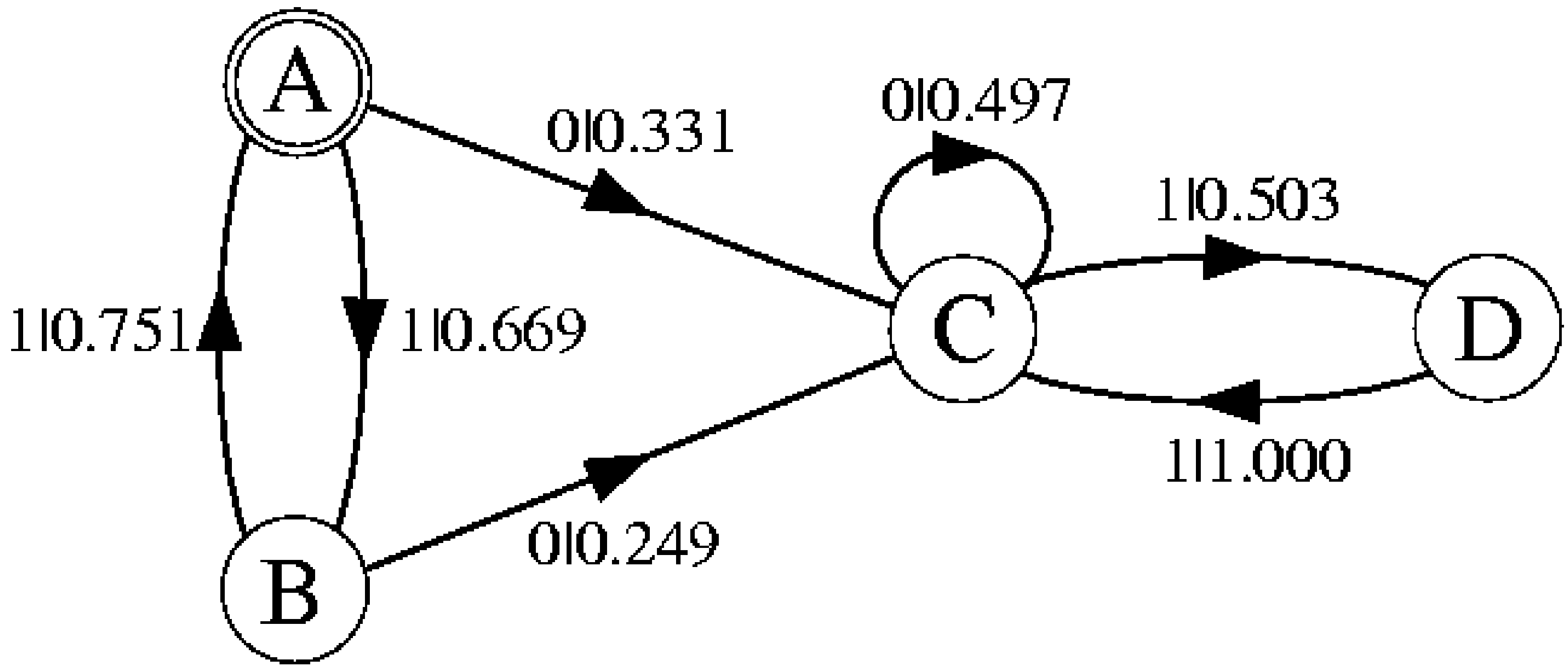
- Error scales like independent samples

$$E[|D(n) - D|] = O(n^{-1/2})$$

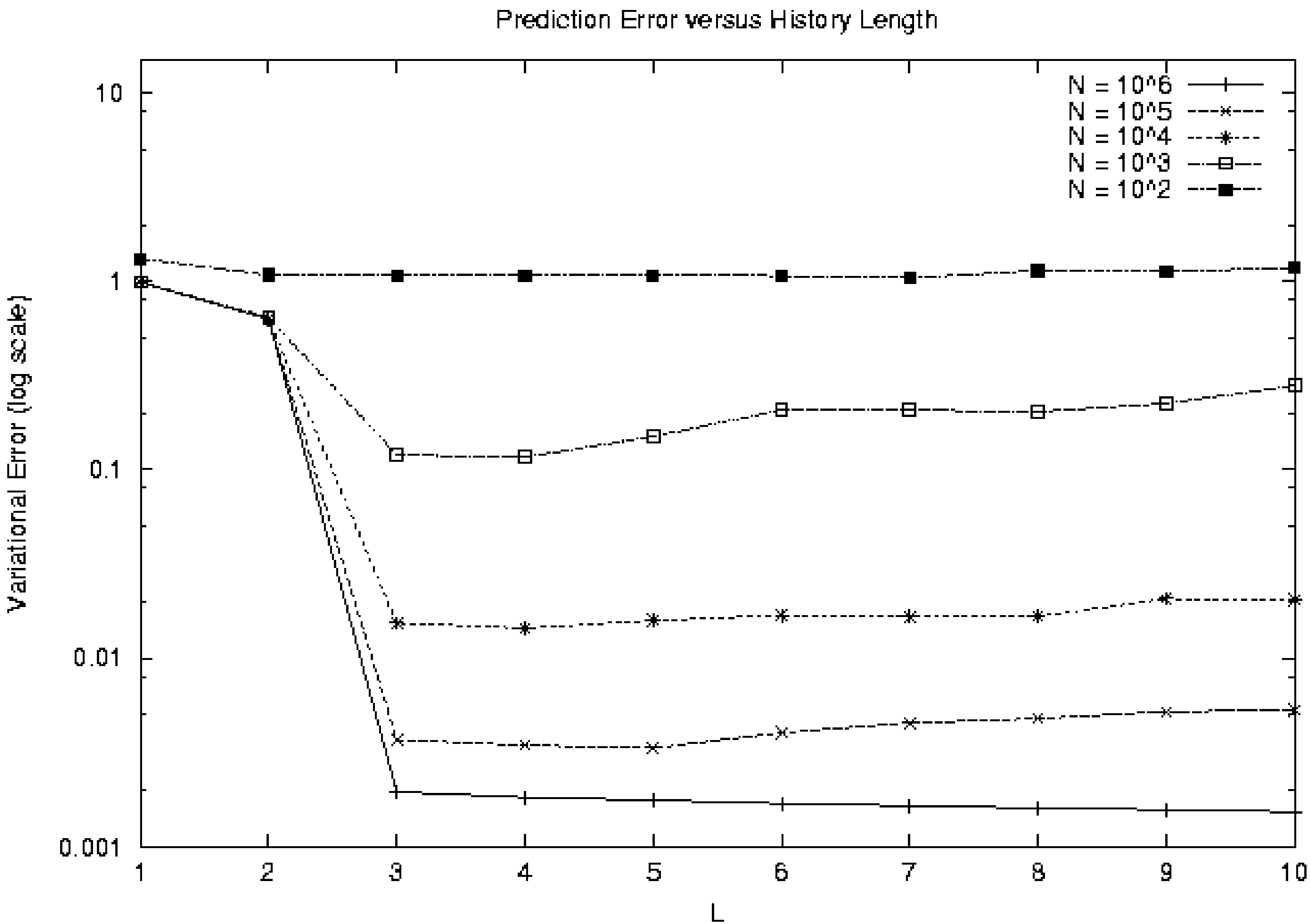
Example

- Even Process
- Language: blocks of 0s, any length, separated by blocks of 1s, even length
- Infinite-range correlation
- Reconstructed with history length 3

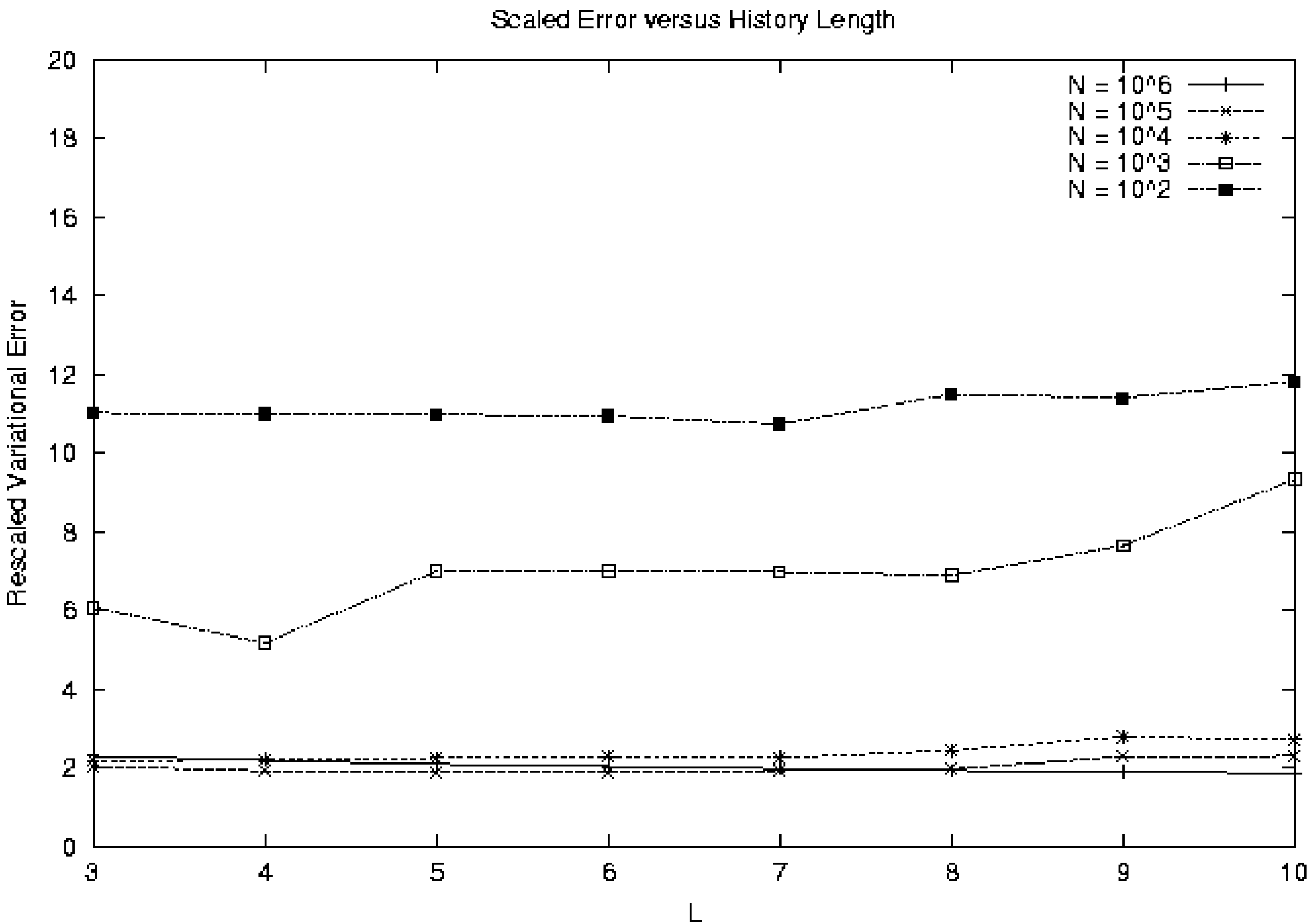




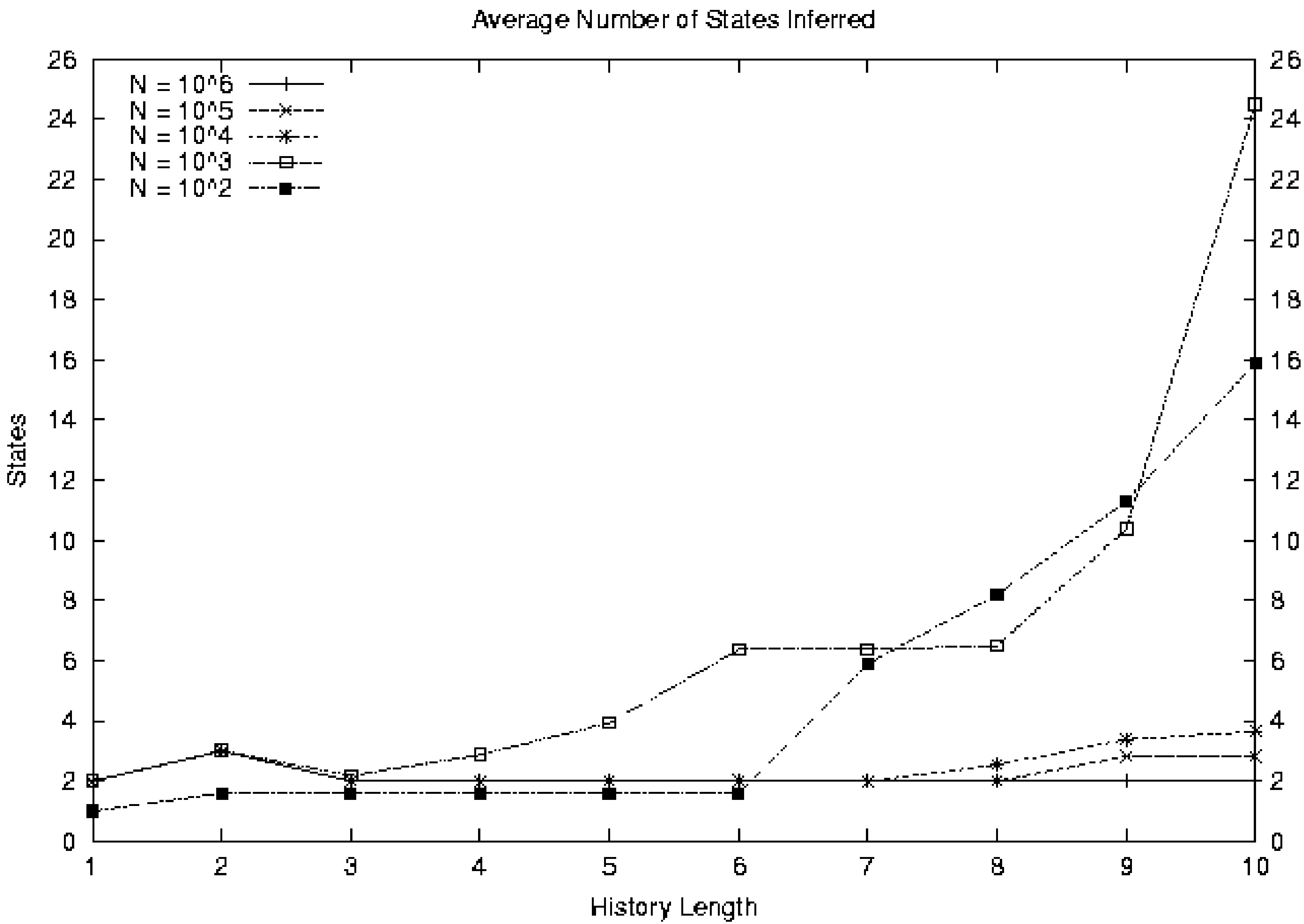
Machine reconstructed from 10,000 symbols and length-3 histories



Prediction Error versus History Length
Variational Error (log scale)
L
N = 10^6
N = 10^5
N = 10^4
N = 10^3
N = 10^2
10
1
0.1
0.01
0.001
1
2
3
4
5
6
7
8
9
10



Scaled Error versus History Length
Rescaled Variational Error
L
N = 10^6
N = 10^5
N = 10^4
N = 10^3
N = 10^2
0
2
4
6
8
10
12
14
16
18
20
3
4
5
6
7
8
9
10


Average Number of States Inferred
N = 10^6
N = 10^5
N = 10^4
N = 10^3
N = 10^2
States
History Length
0
2
4
6
8
10
12
14
16
18
20
22
24
26
1
2
3
4
5
6
7
8
9
10

The Competition

- Variable-length Markov Models, Context Trees, Probabilistic Suffix Trees
- Like CSSR, but always split states so that state ≡ suffix
- Automatically recursive
  - Actually quite old (JANET, 1959)
  - May mimic unconscious human learning
  - Rediscovered in 1980s (Rissanen &c)

- VLMM $\subset$ CSSR
- CSSR $\not\subset$ VLMM
- Even Process
  - State A = *0, *011, *01111, etc.
  - State B = *01, *0111, *011111, etc.
  - VLMM needs ∞ states, CSSR needs 2
- Sofic processes

Extensions

- Transducers, controlled dynamical systems✓
- Continuous-valued series
  - Kernel density estimators?
  - Adaptive discretization, estimating generating partitions?
- Higher-order languages

Information in Networks

with K. L. Shalizi and M. Camperi

- One time series per node
- Do reconstruction on each node separately
- Filter for state series
- Mutual information between states

  = generalized synchrony

  = distributed information
- Architecture via conditional independence?

Spatiotemporal Systems

with R. Haslinger, K. L. Shalizi and J. Usinowicz

- Causal state now local to a point in space and time
- Use forward and reverse light-cones, not time series
- Markov random field, not Markov chain
- “Space” can be arbitrary graph

Take-Home

- Hidden Markov model always available
- Optimal HMM = causal states
- Reconstruction from data