Random Walk Null Models for Time Series

Daryl DeFord

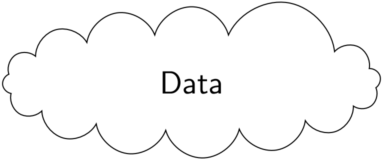
Dartmouth College
Department of Mathematics

Department of Mathematics
Grand Valley State University
Grand Rapids, MI
February 7, 2018

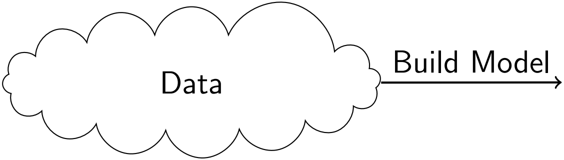
Dartmouth

Outline

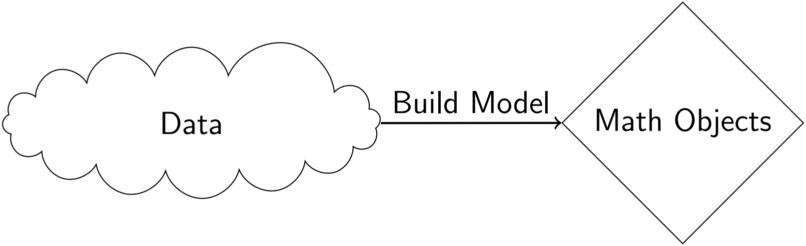
Philosophy



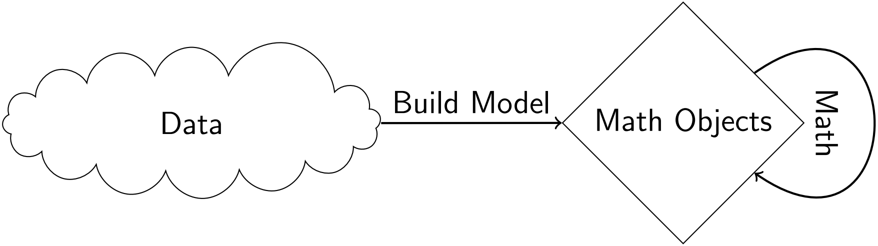
Philosophy



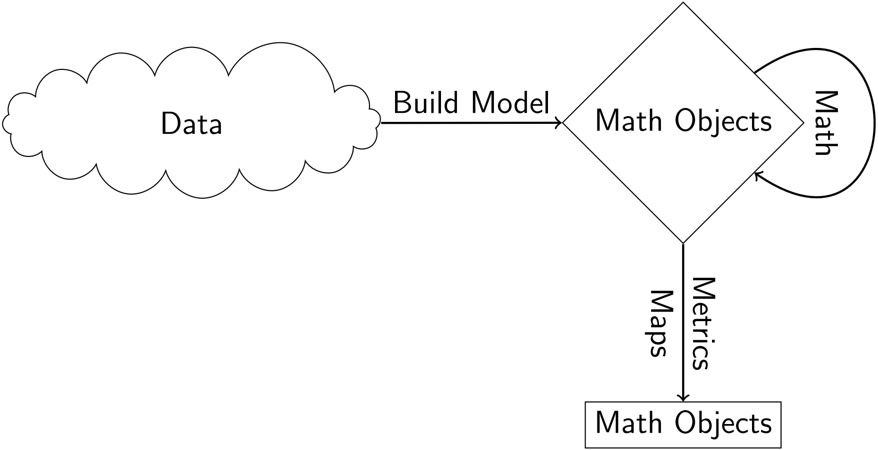
Philosophy



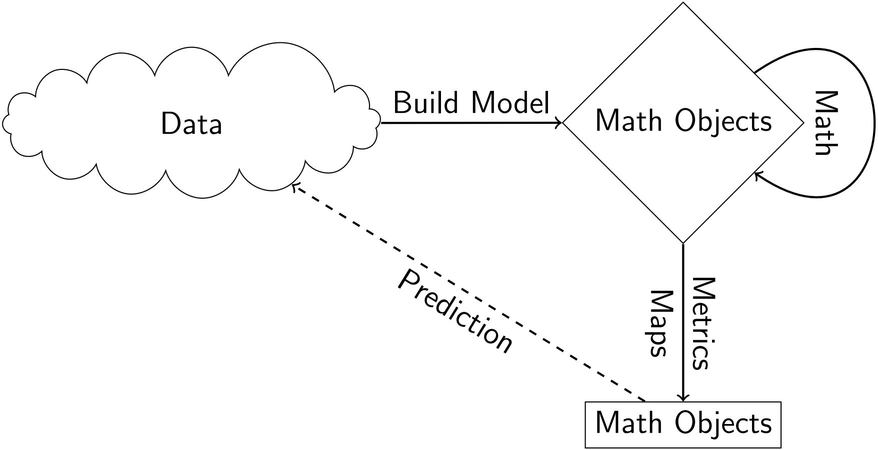
Philosophy



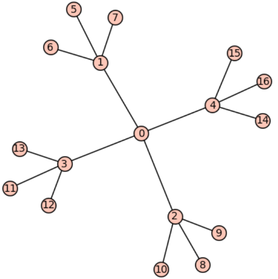
Philosophy


Data
Build Model
Math Objects
Math
Maps
Metrics
Math Objects

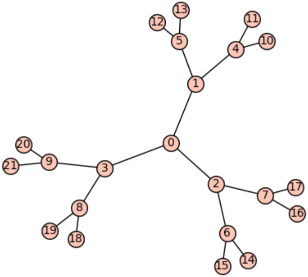
Philosophy


Data
Build Model
Math Objects
Math
Maps
Metrics
Prediction
Math Objects

Adjacency Spectra of Regular Trees



(a) $X_2^{(4)}$



(b) $X_3^{(3)}$

Eigenvalues of (3)–trees

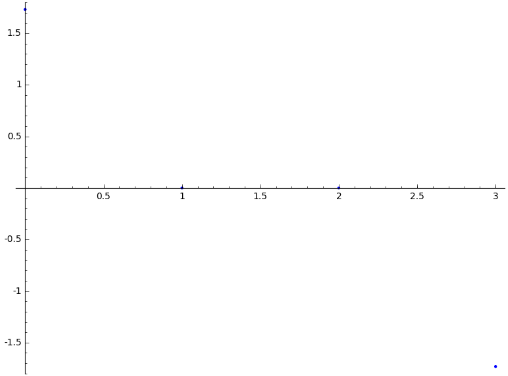


Figure : Eigenvalues of $X_1^{(3)}$

Eigenvalues of (3)–trees

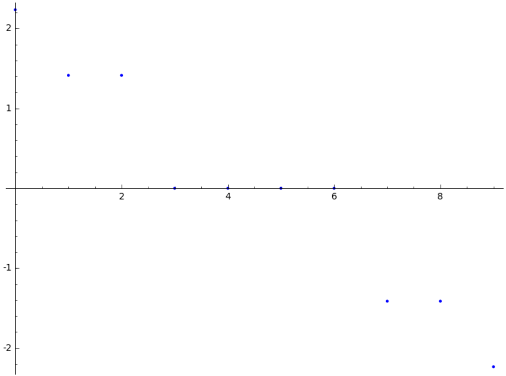


Figure : Eigenvalues of $X_2^{(3)}$

Eigenvalues of (3)–trees

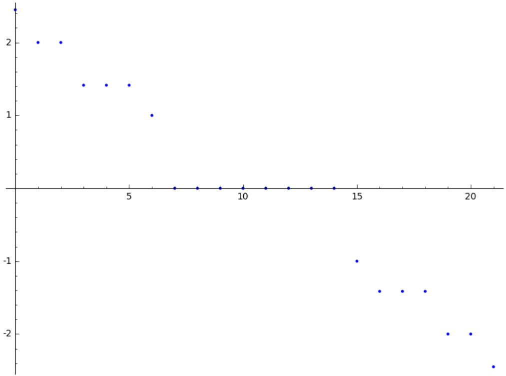


Figure : Eigenvalues of $X_3^{(3)}$

Eigenvalues of (3)–trees

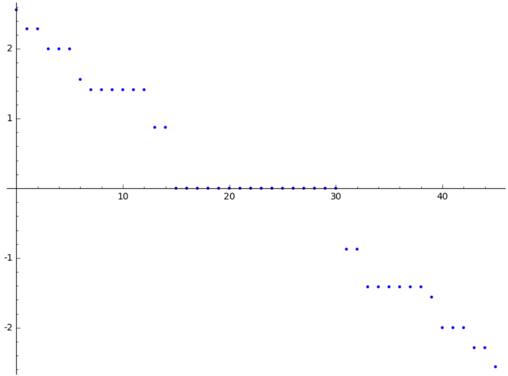


Figure : Eigenvalues of $X_4^{(3)}$

Eigenvalues of (3)–trees

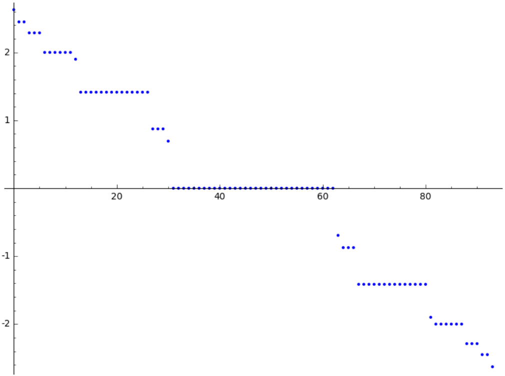


Figure : Eigenvalues of $X_5^{(3)}$

Eigenvalues of (3)–trees

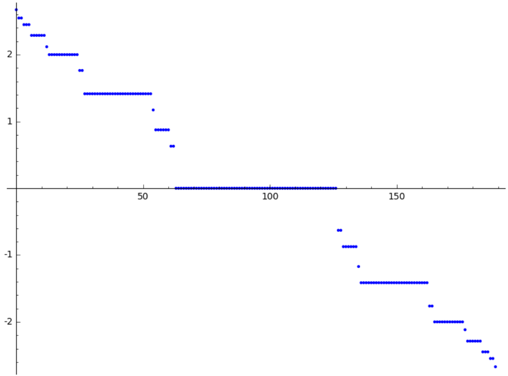


Figure : Eigenvalues of $X_6^{(3)}$

Eigenvalues of (3)–trees

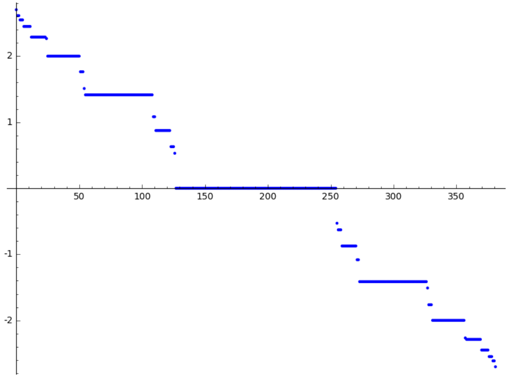


Figure : Eigenvalues of $X_7^{(3)}$

Eigenvalues of (3)–trees

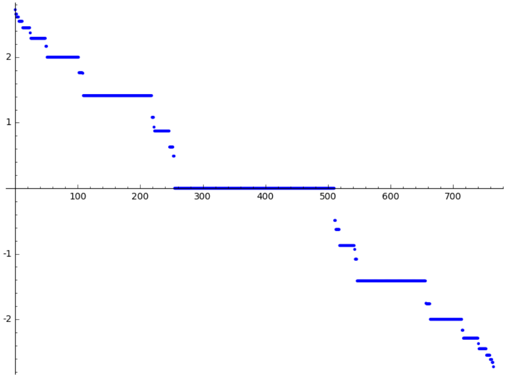


Figure : Eigenvalues of $X_8^{(3)}$

Eigenvalues of Regular Trees

Define two families of polynomials:

$$P_n^k(x) = xP_{n-1}^k - (k-1)P_{n-2}^k$$

with initial conditions $P_0^k(x) = 0$, $P_1^k(x) = 1$, and $P_2^k(x) = x$ and

$$Q_n^k(x) = xP_n^k(x) - kP_{n-1}^k(x).$$

Theorem

The roots of $P_s^k(x)$ for $1 \leq s \leq r$ and $Q_r^k(x)$ are precisely the eigenvalues of the finite k-ary tree $X_r^{(k)}$.

Enumerative Results

Theorem

If λ is a root P_r^k and not a root of P_m^k for any $m < r$ then asymptotically (as $r \longrightarrow \infty$), the proportion of eigenvalues of X_k^r is $\dfrac{(k-2)^2}{(k-1)^r - 1}$.

Corollary

$$\sum_{n=2}^{\infty} \frac{\varphi(n)(k-2)^2}{(k-1)^n - 1} = 1.$$

Corollary ($k = 3$)

$$\sum_{n=1}^{\infty} \frac{\varphi(n)}{2^n - 1} = 2.$$

Tree Questions

Question

Given a graph G and an associated adjacency eigenpair (v, λ) does there exist a subgraph H of G so that $(v|_H, \lambda)$ is an eigenpair for H?

Question

Is there a nice closed form for the endpoints of the Cantor–like sets:

$$\left\{ \sum_{n=1}^{\infty} \frac{(k-2)^2}{(k-1)^n - 1} \sum_{\substack{(\ell,n)=1 \\ \ell < \frac{an}{m}}} 1 : \begin{array}{c} m \in \mathbb{N} \\ (a,m) = 1 \end{array} \right\}$$

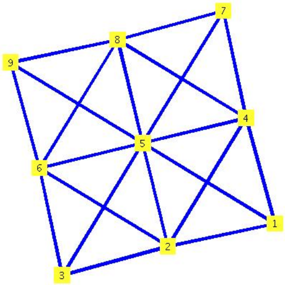
Question

Can we characterize the sequences of graphs $G_1, G_2, \ldots$ satisfying for all $\varepsilon > 0$ there exists a finite set $\Lambda \subset \mathbb{R}$ and a $N \in \mathbb{N}$ such that for all $n > N$:

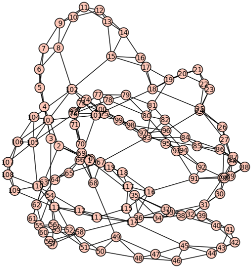
$$\frac{|\{\lambda \in \operatorname{spec}(G_n) : \lambda \notin \Lambda\}|}{|\operatorname{spec}(G_n)|} < \varepsilon.$$

uth

Network Example



(a) Graph

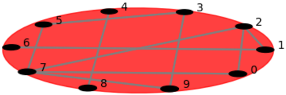


(b) Network

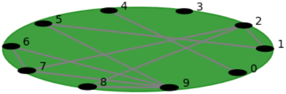
Multiplex Networks

Definition

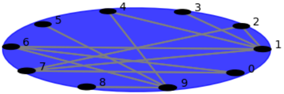
A *multiplex* is a collection of graphs all defined on the same node set.



(a) Family

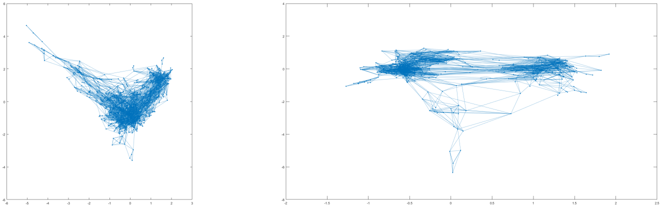


(b) Colleagues



(c) Facebook

Karnataka Village Data[1]



(a) Village 4 (b) Village 61

Figure : Karnataka Villages

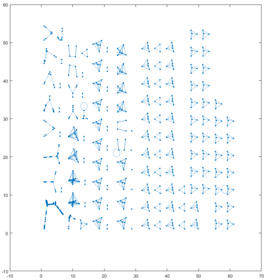
[1] A. Banerjee, A.G. Chandrasekhar, E. Duflo, and M.O. Jackson, The Diffusion of Microfinance. Science, (2013).

Village Layers

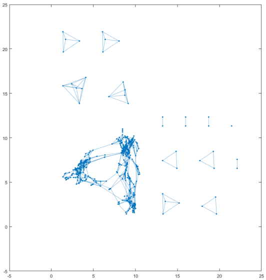
Layer	Village 4			Village 61		
Description	Density	Comp.	Giant %	Density	Comp.	Giant %
Borrow Money	.0082	26	.8354	.0108	15	.9188
Give Advice	.0077	49	.5892	.0098	34	.7377
Help Make Decisions	.0076	61	.1277	.0100	24	.8562
Borrow Kerosene or Rice	.0085	21	.8338	.0113	14	.9171
Lend Kerosene or Rice	.0086	22	.8308	.0113	14	.9255
Lend Money	.0081	14	.7908	.0107	17	.9036
Medical Advice	.0075	84	.2938	.0106	14	.9306
Friends	.0089	15	.9277	.0105	22	.8714
Relatives	.0085	29	.7231	.0105	26	.5448
Attend Temple With	.0073	117	.0462	.0089	108	.0372
Visit Their Home	.0087	15	.9185	.0116	11	.9475
Visit Your Home	.0088	16	.9108	.0117	11	.9492
Aggregate	.0121	3	.9862	.0155	8	.9679

Table : Layer information for two of the Karnataka Villages.

Medical Advice

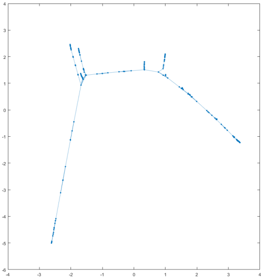


(a) Village 4

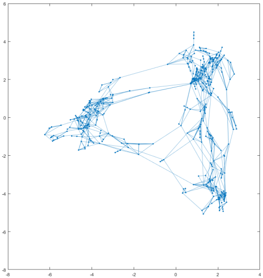


(b) Village 61

Medical Advice



(a) Village 4



(b) Village 61

Multiplex Questions

Question (How to account for layer heterogeneity?)

- *Structural Consequences*
- *Dynamical Consequences*
- *Spectral Consequences*

Question (How to account for node indivisibility?)

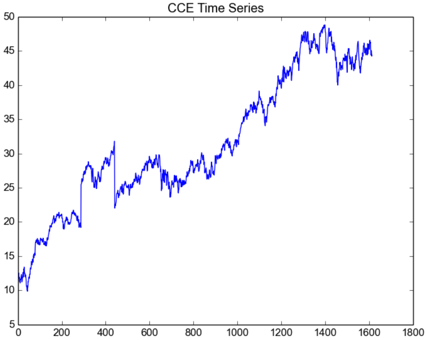
- *Merge Centrality Measures*
- *Define Neighborhoods*
- *Multi–Membership Communities*

Question (Data Examples)

- *Social Networks*
- *Transportation Networks*
- *Economic Networks*
- *Paired Networks*

uth

Time Series


CCE Time Series
50
45
40
35
30
25
20
15
10
5
0
200
400
600
800
1000
1200
1400
1600
1800

Iterated Maps

Given a function $f : [0, 1] \to [0, 1]$ and a point $x \in [0, 1]$, consider the behavior of $\{x, f(x), f(f(x)), f(f(f(x))), \ldots\}$.

Iterated Maps

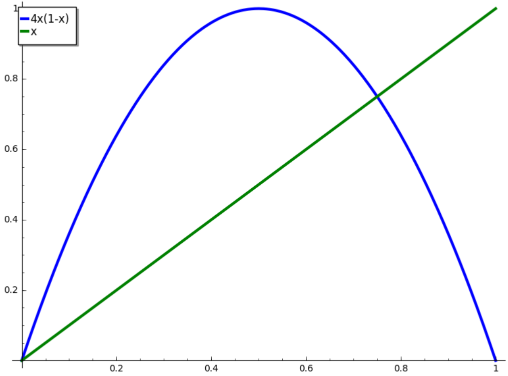
Given a function $f : [0,1] \to [0,1]$ and a point $x \in [0,1]$, consider the behavior of $\{x, f(x), f(f(x)), f(f(f(x))), \ldots\}$.

Example

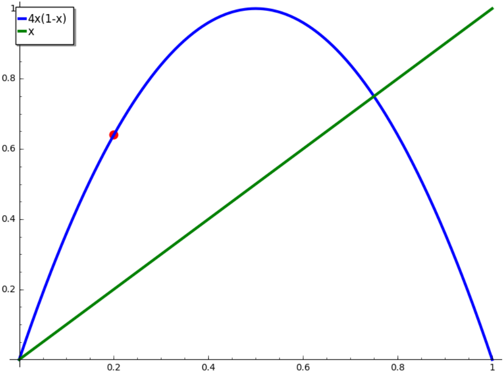
Let $f(x) = 4x(1-x)$ *and* $x_0 = .2$. *Then, the list of values is:*

$$[0.20, 0.64, 0.92, 0.28, 0.82, 0.58, 0.97, 0.11, 0.40, \ldots].$$

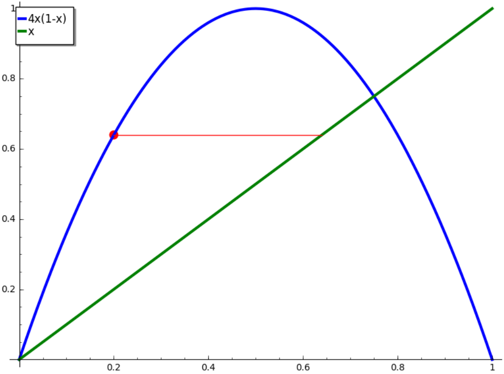
Iterated Example



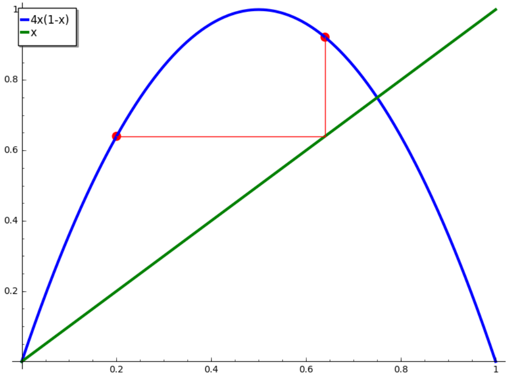
Iterated Example



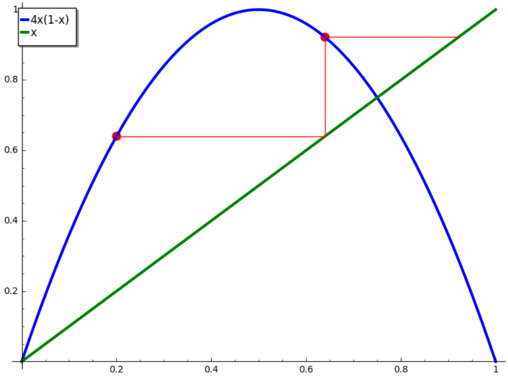
Iterated Example



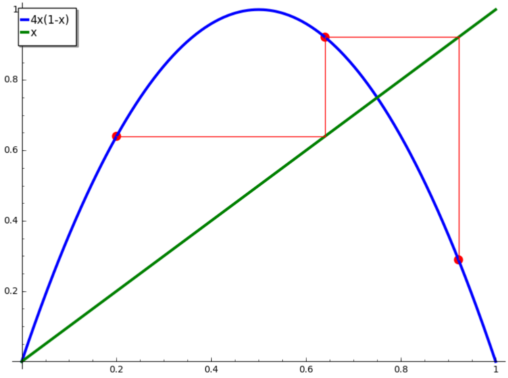
Iterated Example (12)



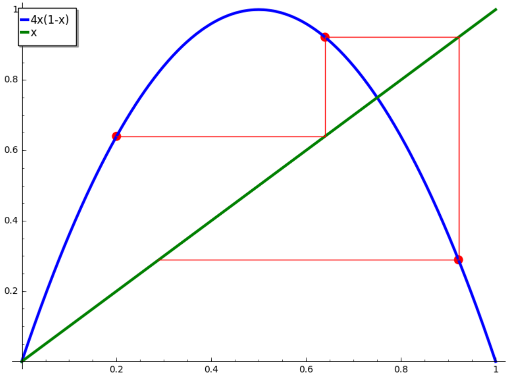
Iterated Example (12)



Iterated Example (231)

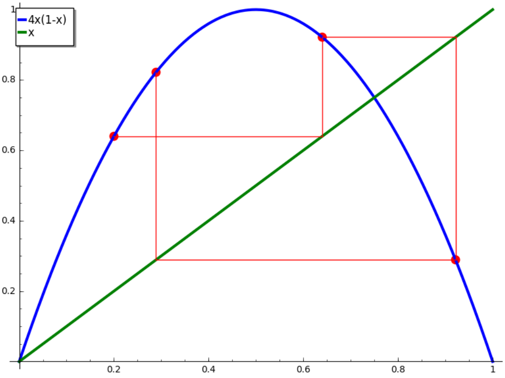


Iterated Example (231)

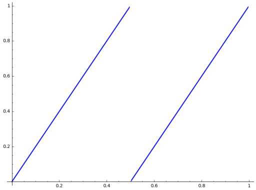


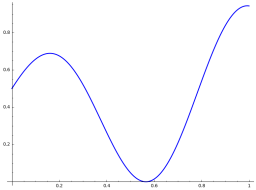
Dartmouth

Iterated Example (2413)

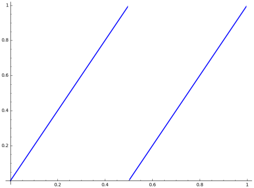


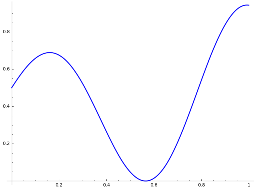
Forbidden Patterns





Forbidden Patterns

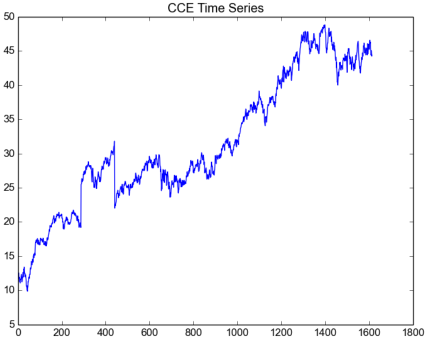




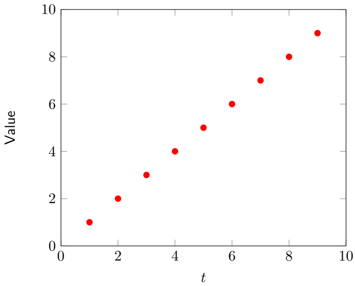
Definition (Topological Entropy)

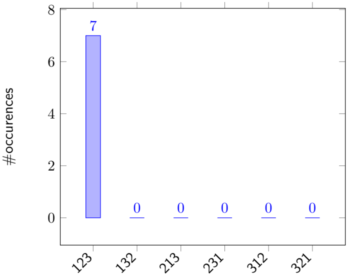
$$TE = \lim_{n\to\infty} \frac{\log(|\mathrm{Allow}_n(f)|)}{n-1}$$

Time Series


CCE Time Series
50
45
40
35
30
25
20
15
10
5
0
200
400
600
800
1000
1200
1400
1600
1800

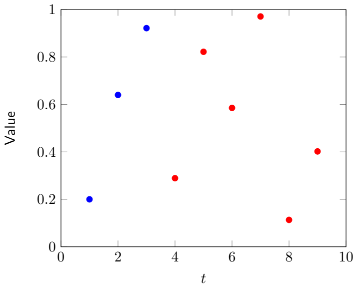
Simple Time Series

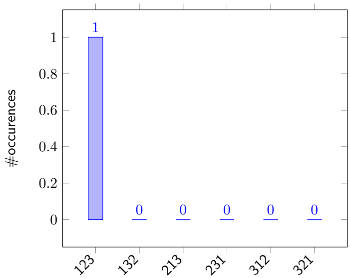




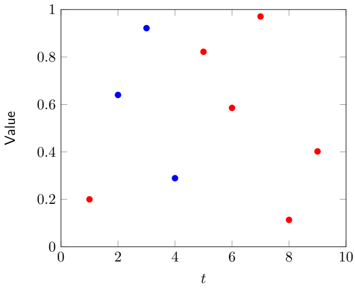
Dartmouth

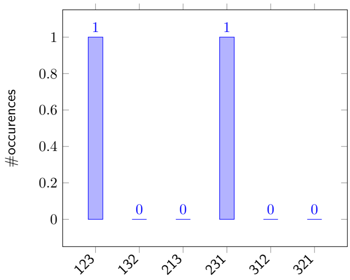
Complex Time Series



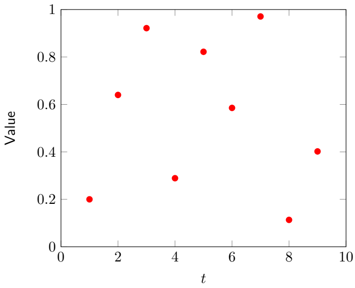


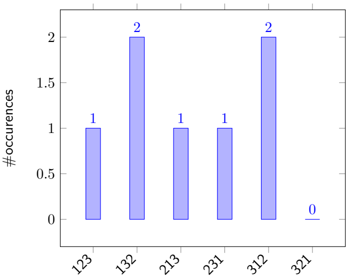
Complex Time Series





Complex Time Series





Dartmouth

Complexity Measures

Definition (Normalized Permutation Entropy)

$$NPE(\{X_i\}) = -\frac{1}{\log(N!)} \sum_{\pi \in S_n} p_\pi \log(p_\pi)$$

Complexity Measures

Definition (Normalized Permutation Entropy)

$$NPE(\{X_i\}) = -\frac{1}{\log(N!)} \sum_{\pi \in S_n} p_\pi \log(p_\pi)$$

Definition (Uniform KL Divergence)

$$D_{KL}(\{X_i\}||\text{UNIFORM}) = \sum_{\pi \in S_n} p_\pi \log\left(\frac{p_\pi}{\frac{1}{n!}}\right)$$

Complexity Measures

Definition (Normalized Permutation Entropy)

$$NPE(\{X_i\}) = -\frac{1}{\log(N!)} \sum_{\pi \in S_n} p_\pi \log(p_\pi)$$

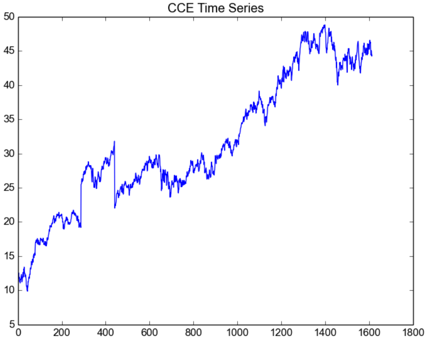
Definition (Uniform KL Divergence)

$$D_{KL}(\{X_i\}||\text{UNIFORM}) = \sum_{\pi \in S_n} p_\pi \log\left(\frac{p_\pi}{\frac{1}{n!}}\right)$$

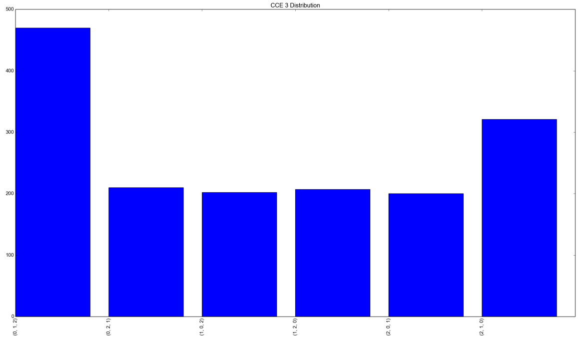
Observation

$$1 - NPE(\{X_i\}) = \frac{1}{\log(N!)} D_{KL}(\{X_i\}||\text{UNIFORM})$$

Stock Data (Closing Prices)

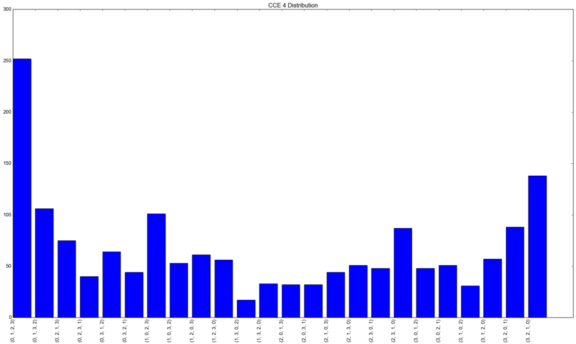


Stock Data (n=3)

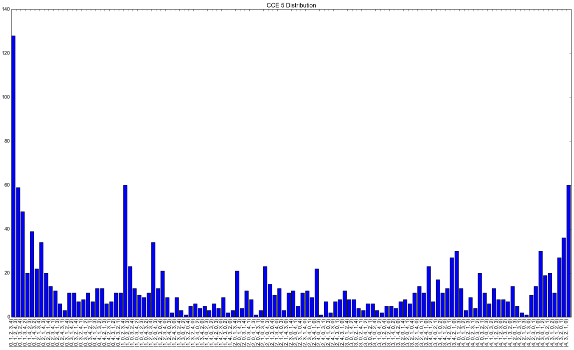


Dartmouth

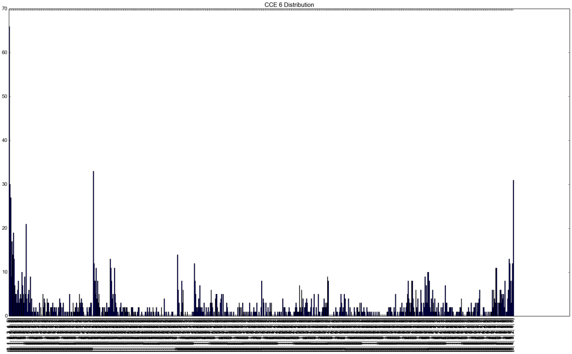
Stock Data (n=4)



Stock Data (n=5)



Stock Data (n=6)



Dartmouth

Random Walk Null Models

Definition (Random Walk)

Let $\{X_i\}$ be a set of I.I.D. continuous random variables and define the "random walk" $\{Z_i\}$ by $Z_j = \sum_{i=0}^{j} X_j$.

Random Walk Null Models

Definition (Random Walk)

Let $\{X_i\}$ be a set of I.I.D. continuous random variables and define the "random walk" $\{Z_i\}$ by $Z_j = \sum_{i=0}^{j} X_j$.

Proposition (No Forbidden Patterns)

If $\{Z_i\}$ are defined as above then every permutation occurs with some positive probability.

Random Walk Null Models

Definition (Random Walk)

Let $\{X_i\}$ be a set of I.I.D. continuous random variables and define the "random walk" $\{Z_i\}$ by $Z_j = \sum_{i=0}^{j} X_j$.

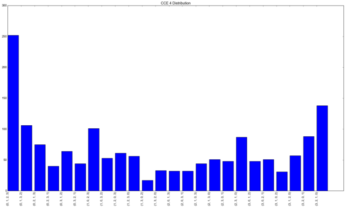
Proposition (No Forbidden Patterns)

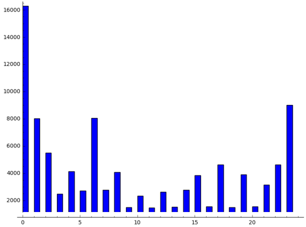
If $\{Z_i\}$ are defined as above then every permutation occurs with some positive probability.

Proposition (No Uniform Distribution)

If $\{Z_i\}$ are defined as above and $n \geq 3$ then the expected distribution of permutations is not uniform.

Uniform Steps CCE





New Complexity Measure

Definition (Null Model KL Divergence)

$$\mathsf{D_{KL}}_n(X) := \mathsf{D_{KL}}_n(X||Z) = \sum_{\pi \in \mathcal{S}_n} p_\pi \log\left(\frac{p_\pi}{q_\pi}\right),$$

where p_π is the relative frequency of π in X and q_π is the relative frequency of π in Z.

Hyperplanes

Example (Uniformly distributed steps)

In order for the pattern 1342 *to appear in the random walk time series we need the following inequalities to hold:*

- $X_1 > 0$
- $X_2 > 0$
- $X_3 < 0$
- $X_3 > X_2$
- $X_3 < X_1 + X_2$

Integration Regions

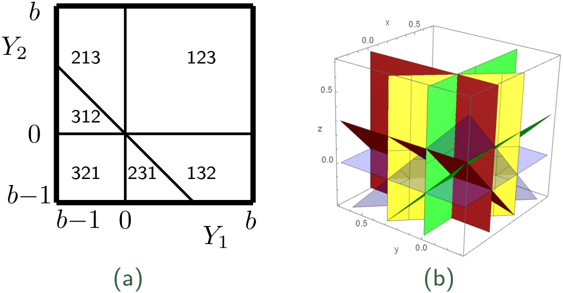


Figure : The regions of integration for patterns in uniform random walks for (**a**) $n = 3$ and (**b**) $n = 4$, sketched here for $b = 0.65$.

Null Distributions ($n = 3$)

Pattern	Normal: $\mu = 0$	Uniform: $\mu = 0$	Uniform: $\mathbb{P}(Y > 0) = b$
{123}	$1/4$	$1/4$	b^2
{132, 213}	$1/8$	$1/8$	$(1/2)(1-b)^2$
{231, 312}	$1/8$	$1/8$	$(1/2)(b^2+2b-1)$
{321}	$1/4$	$1/4$	$(1-b)^2$

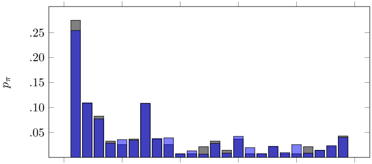
Dartmouth

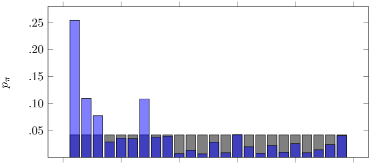
Null Distributions ($n = 4$)

Pattern	Normal: $\mu = 0$	Uniform: $\mu = 0$	Uniform: $\mathbb{P}(Y > 0) = b$
{1234}	0.1250	1/8	b^3
{1243, 2134}	0.0625	1/16	$(1/2)b(1-b)(3b-1)$
{1324}	0.0417	1/24	$(1/3)(1-b)(7b^2-5b+1)$
{1342, 3124}	0.0208	1/24	$(1/6)(1-b)^2(4b-1)$
{1423, 2314}	0.0355	1/48	$(1/6)(1-b)^2(5b-2)$
{1432, 2143, 3214}	0.0270	1/48	$\begin{cases}(1/6)(2-24b+48b^2-15b^3) & \text{if } b \le 2/3\\ (b-1)^2(2b-1) & \text{if } b > 2/3\end{cases}$
{2341, 3412, 4123}	0.0270	1/48	$(1/6)(1-b)^3$
{2413}	0.0146	1/48	$(1/6)(1-b)^3$
{2431, 4213}	0.0208	1/24	$\begin{cases}(1/6)(24b^3-45b^2+27b-5) & \text{if } b \le 2/3\\ (1/2)(1-b)^3 & \text{if } b > 2/3\end{cases}$
{3142}	0.0146	1/48	$\begin{cases}(1/6)(25b^3-48b^2+30b-6) & \text{if } b \le 2/3\\ (1/3)(1-b)^3 & \text{if } b > 2/3\end{cases}$
{3241, 4132}	0.0355	1/48	$(1/6)(1-b)^3$
{3421, 4312}	0.0625	1/16	$(1/2)(1-b)^3$
{4231}	0.0417	1/24	$(1/3)(1-b)^3$
{4321}	0.1250	1/8	$(1-b)^3$

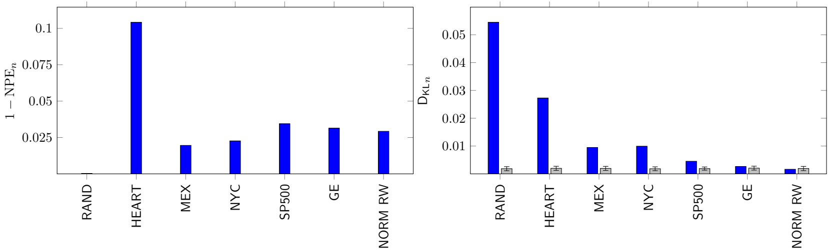
Dartmouth

Uniform Steps S&P 500



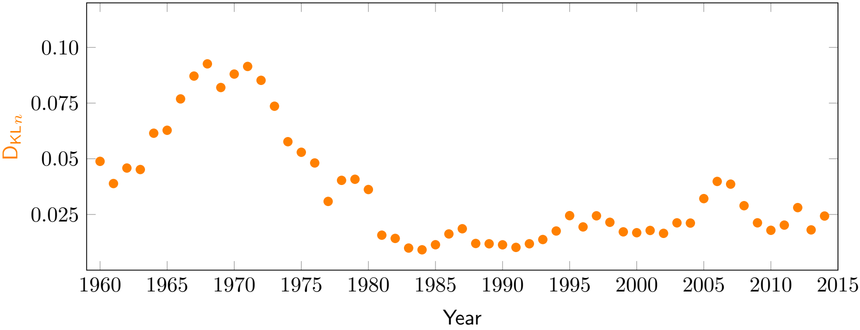


Data Comparisons



(a) (b)

Stock Market Example



Time Series Questions

Question (Random Walk Applications)

- *Can we identify other interesting economic events?*
- *Medical Data?*
- *Climate Data?*

Question (Periodic Null Models)

- *What is the right null model or probability distribution for periodic data?*
- *Higher order Markov models?*
- *Iterated functions plus noise?*

Question (Walks on S_n)

How does the steady state of the Markov process on the patterns themselves compare to the probabilistic null models?

uth

Time Series References

- C. Bandt: *Permutation Entropy and Order Patterns in Long Time Series*, Time Series Analysis and Forecsting, Springer, 2016.
- D. DeFord and K. Moore: *Random Walk Null Models for Time Series Data*, Entropy, 19(11), 615, 2017.
- M. Zanin: *Forbidden patterns in financial time series*, Chaos 18 (2008) 013119.
- M. Zanin, L. Zunino, O.A. Rosso, and D. Papo : *Permutation Entropy and Its Main Biomedical and Econophysics Applications: A Review*, Entropy 2012, 14, 1553-1577.

That's all..

Thank You!