Permutation Complexity Measures for Time Series

Daryl DeFord and Kate Moore

Dartmouth College
Department of Mathematics

SIAM Annual Meeting
Pittsburgh, PA
July 11, 2017

Dartmouth

Abstract

Permutation entropy has become a standard tool in time–series analysis that exploits the the temporal properties of these data sets. Many current applications use an approach based on Shannon entropy, which implicitly assumes an underlying uniform distribution on patterns. In this paper, we consider several additional null models for time series data and determine the corresponding permutation distributions. This allows us to compare real–world data to more complex generative processes. Additionally, building on recent results of Martinez, we define a measure of complexity that allows us to characterize when a random walk is an appropriate model for a time series.

Outline

Permutation Patterns

$$\begin{pmatrix} 3 & 6 & 7 & 9 & 4 & 8 & 1 & 2 & 5 \\ 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 \end{pmatrix}$$

Permutation Patterns (123)

$$\begin{pmatrix} \color{blue}{3} & \color{blue}{6} & \color{blue}{7} & 9 & 4 & 8 & 1 & 2 & 5 \\ 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 \end{pmatrix}$$

Permutation Patterns (123)

$$\begin{pmatrix} 3 & \color{blue}{6} & \color{blue}{7} & \color{blue}{9} & 4 & 8 & 1 & 2 & 5 \\ 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 \end{pmatrix}$$

Permutation Patterns (231)

$$\begin{pmatrix} 3 & 6 & \color{blue}{7} & \color{blue}{9} & \color{blue}{4} & 8 & 1 & 2 & 5 \\ 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 \end{pmatrix}$$

Permutation Patterns (312)

$$\begin{pmatrix} 3 & 6 & 7 & \color{blue}{9} & \color{blue}{4} & \color{blue}{8} & 1 & 2 & 5 \\ 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 \end{pmatrix}$$

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

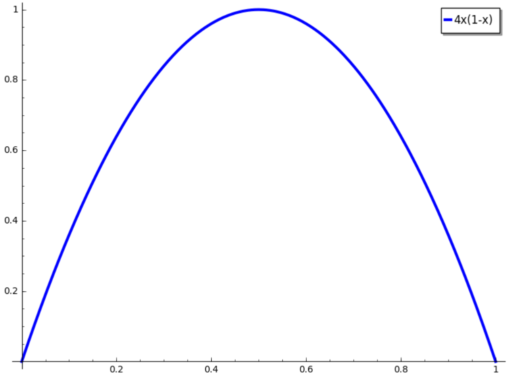
Given a function $f : [0,1] \to [0,1]$ and a point $x \in [0,1]$, consider the behavior of $\{x, f(x), f(f(x)), f(f(f(x))), \ldots\}$.

Example

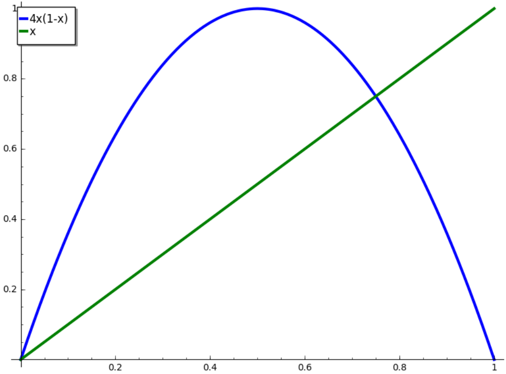
Let $f(x) = 4x(1-x)$ and $x_0 = .2$. Then, the list of values is:

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

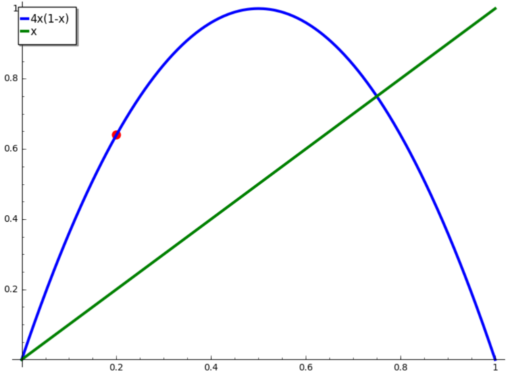
Iterated Example



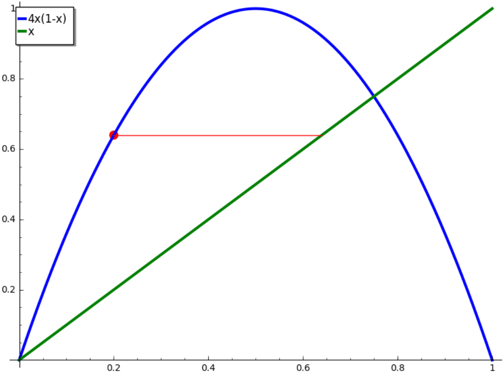
Iterated Example



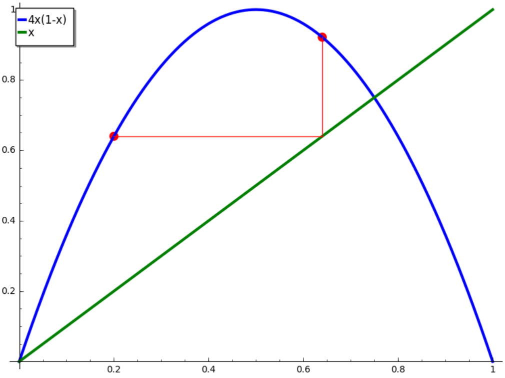
Iterated Example



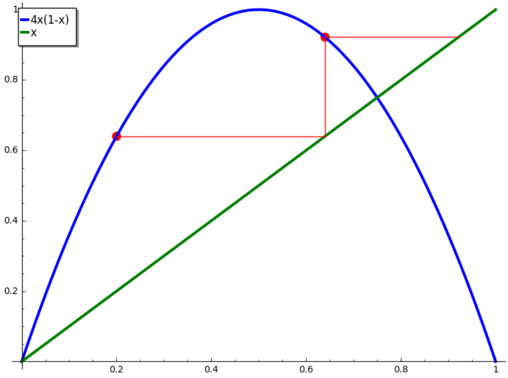
Iterated Example



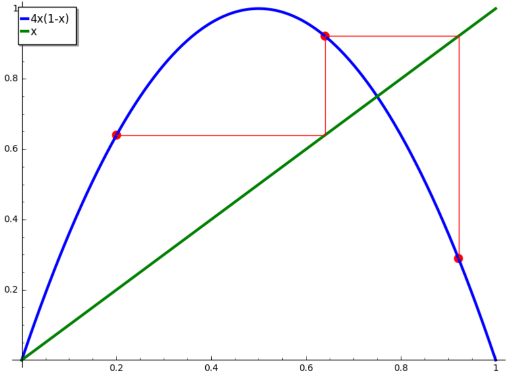
Iterated Example (12)



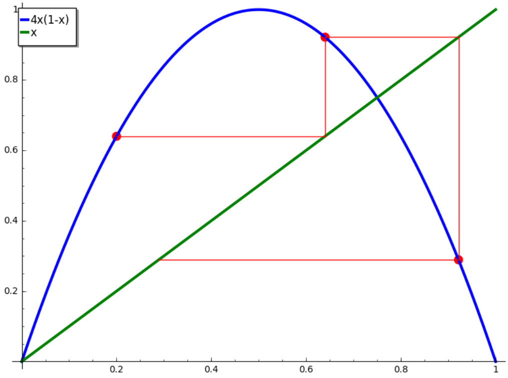
Iterated Example (12)



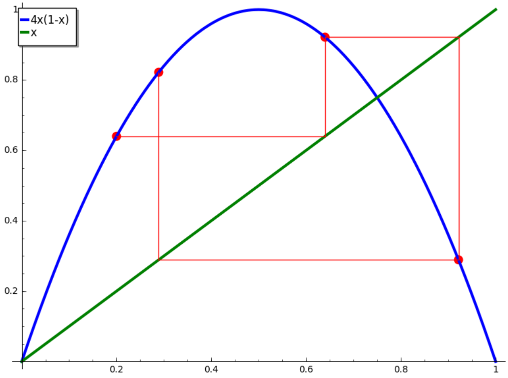
Iterated Example (231)



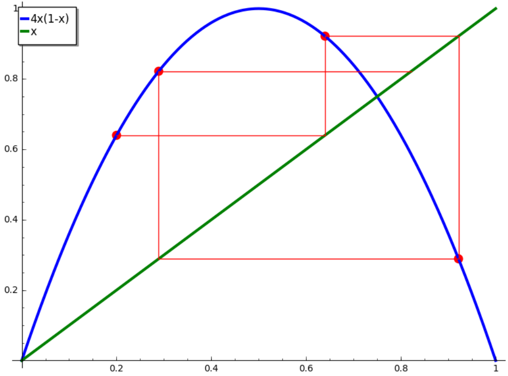
Iterated Example (231)



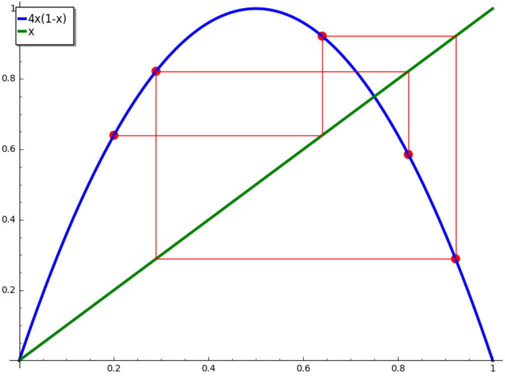
Iterated Example (2413)



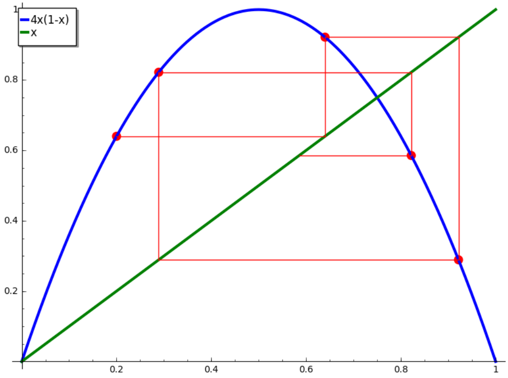
Iterated Example (2413)



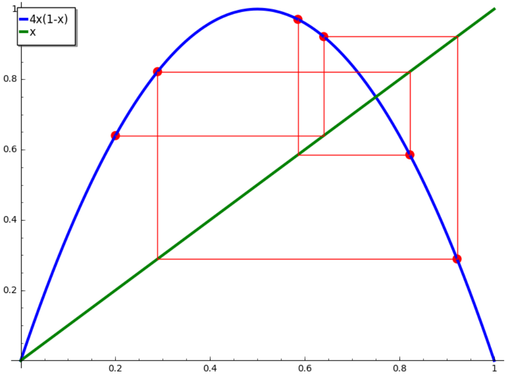
Iterated Example (35142)



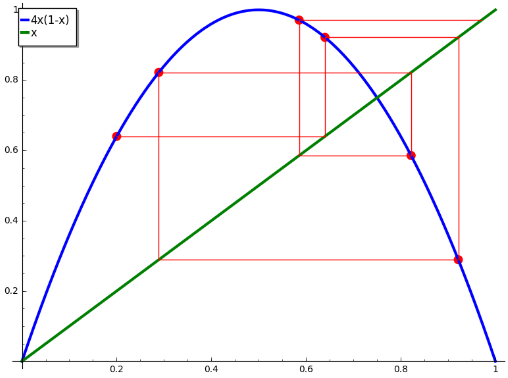
Iterated Example (35142)



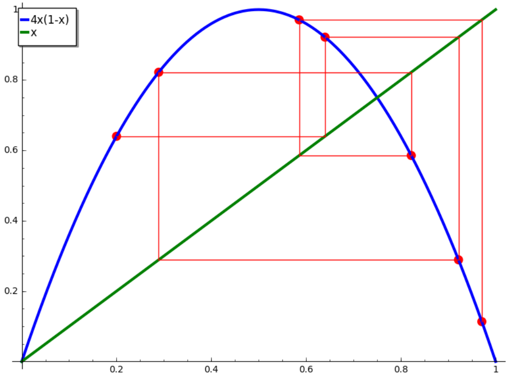
Iterated Example (351426)



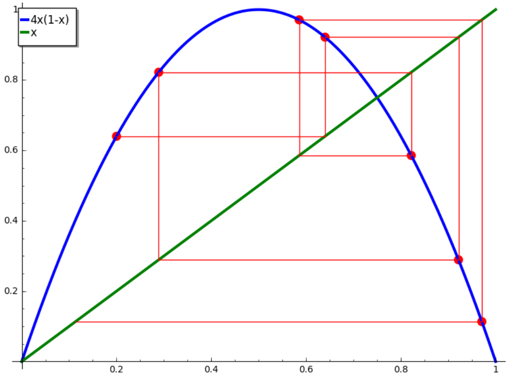
Iterated Example (351426)



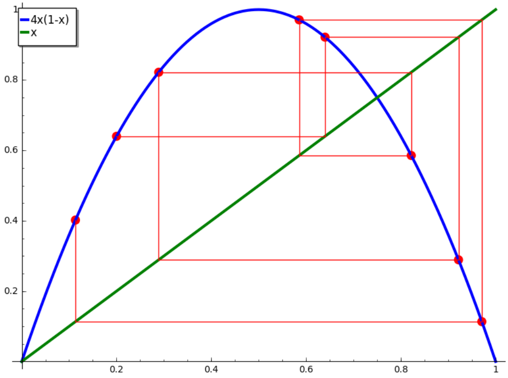
Iterated Example (4625371)



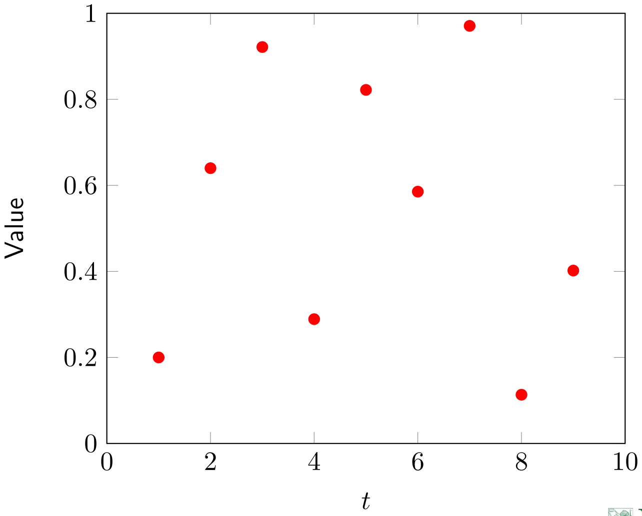
Iterated Example (4625371)



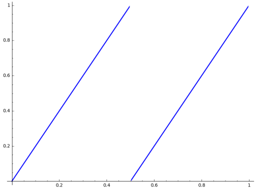
Iterated Example (57264813)

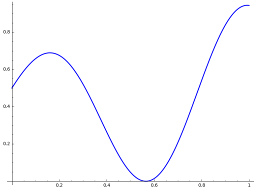


Iterated Example (268375914)

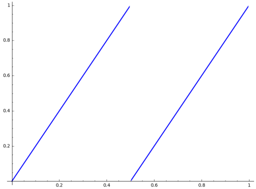


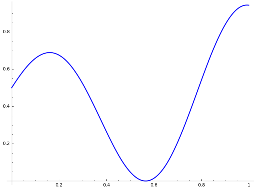
Forbidden Patterns





Forbidden Patterns

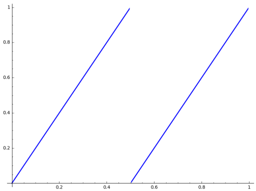


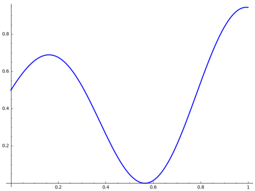


Definition (Topological Entropy)

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

Forbidden Patterns





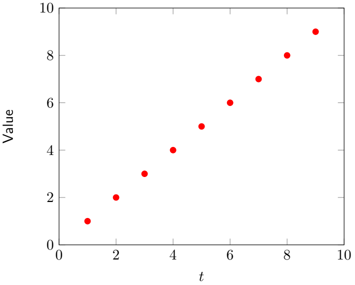
Definition (Topological Entropy)

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

- C. Bandt and B. Pompe: *Permutation entropy: A natural complexity measure for time series*, Phys. Rev. Lett. 88, 174102 (2002).
- C. Bandt, G. Keller, and B. Pompe: *Entropy of interval map s via permutations*, Nonlinearity 15, 1595 (2002).

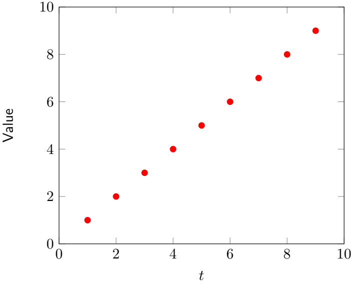
Dartmouth

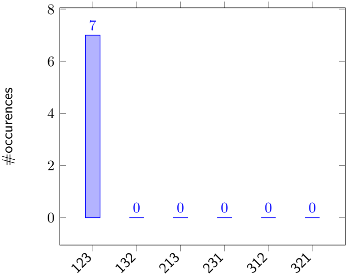
Simple Time Series



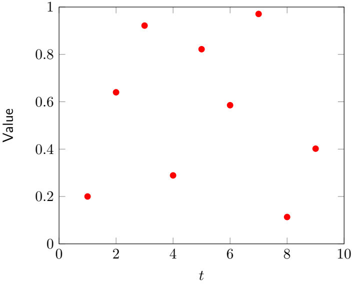
Dartmouth

Simple Time Series



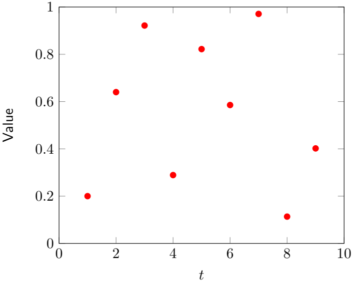


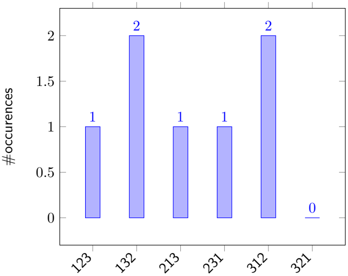
Complex Time Series



Dartmouth

Complex Time Series





Time Series Complexity

Definition (Permutation Entropy)

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

Time Series Complexity

Definition (Permutation Entropy)

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

Definition (Uniform KL Divergence)

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

Time Series Complexity

Definition (Permutation Entropy)

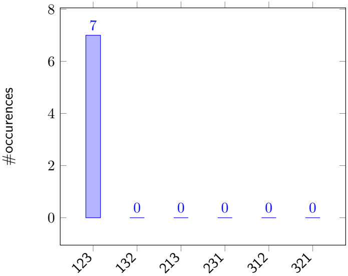
$$PE(\{X_i\}) = -\frac{1}{\log(N!)} \sum_{\pi \in S_n} p_\pi \log(p_\pi)$$

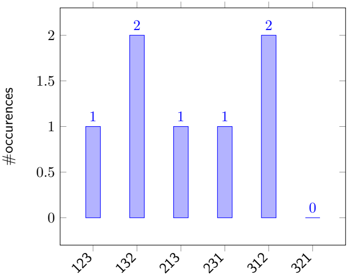
Definition (Uniform KL Divergence)

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

- 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.
- M. Zanin: *Forbidden patterns in financial time series*, Chaos 18 (2008) 013119.
- C. Bandt: *Permutation Entropy and Order Patterns in Long Time Series*, Time Series Analysis and Forecsting, Springer, 2016.

Histogram Comparison





I.I.D. Variables

The uniformity assumption is already associated to a null model.

Lemma

Let $\{X_i\}$ *be a set of I.I.D. random variables, then* (π, t) *and* (τ, s) *are independent random variables if and only if* $|t - s| > n$.

I.I.D. Variables

The uniformity assumption is already associated to a null model.

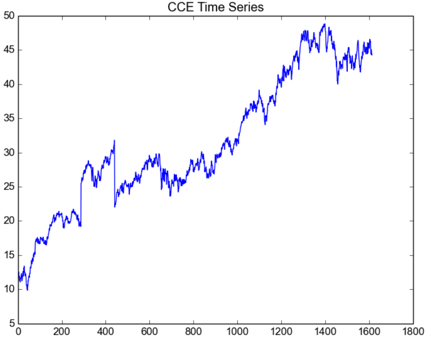
Lemma

Let $\{X_i\}$ be a set of I.I.D. random variables, then (π, t) and (τ, s) are independent random variables if and only if $|t - s| > n$.

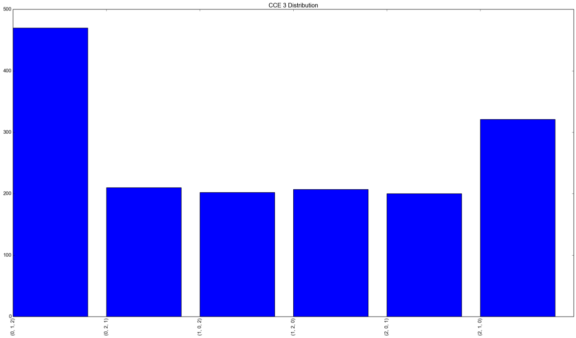
Lemma

Let $\{X_i\}$ be a set of I.I.D. random variables, then $\mathbb{P}(\pi) = \frac{1}{n!}$ for all $\pi \in S_n$.

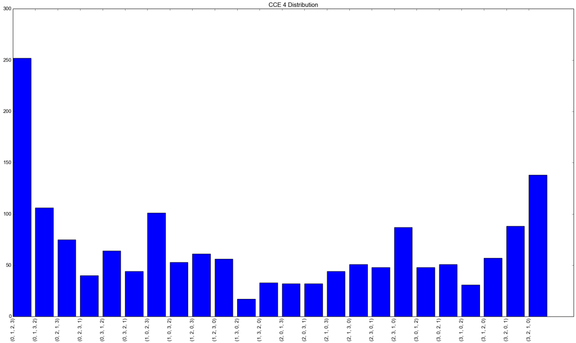
Stock Data (Closing Prices)



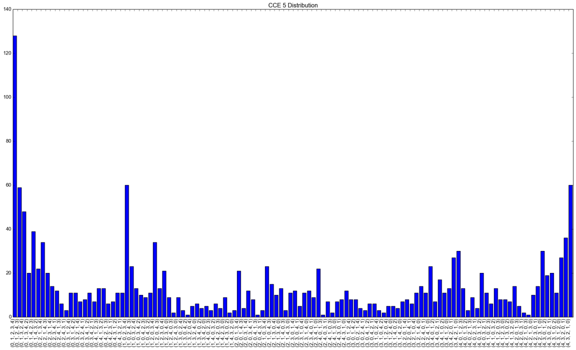
Stock Data (n=3)



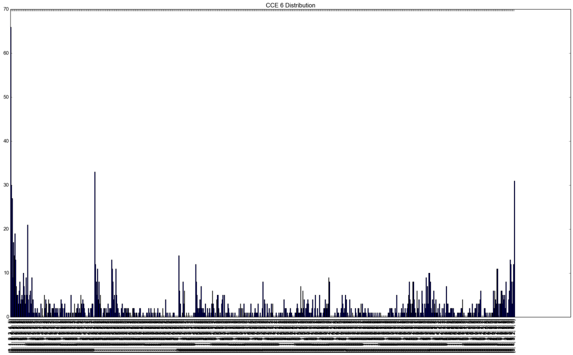
Stock Data (n=4)



Stock Data (n=5)



Stock Data (n=6)



Random Walk Models

Definition (Random Walk)

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

Random Walk Models

Definition (Random Walk)

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

Theorem

If $\{Z_i\}$ are defined as above then either $123\ldots n$ or $n(n-1)(n-2)\ldots 1$ occurs with the highest probability.

Random Walk Models

Definition (Random Walk)

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

Theorem

If $\{Z_i\}$ are defined as above then either $123\ldots n$ or $n(n-1)(n-2)\ldots 1$ occurs with the highest probability.

Corollary

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

Uniform Walks

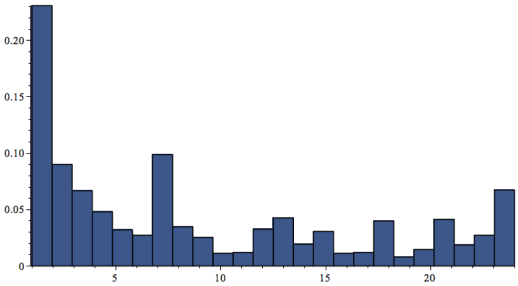
Definition

Let the $\{X_i\}$ be defined as uniform random variables over $[b-1, b]$ where $0 < b < 1$ and define $\{Z_i\}$ with $Z_j = \sum_{i=0}^{j} X_i$.

Uniform Walks

Definition

Let the $\{X_i\}$ be defined as uniform random variables over $[b-1, b]$ where $0 < b < 1$ and define $\{Z_i\}$ with $Z_j = \sum_{i=0}^{j} X_i$.



Expected Distributions

Theorem

Let $\{Z_i\}$ be a random walk as defined above, then the expected distributions of permutations for S_3 and S_4 are characterized by the following:

$\mathbb{P}(123) = b^2$ $\mathbb{P}(132) + \mathbb{P}(231) = (1-b)b$	$\mathbb{P}(321) = (1-b)^2$ $\mathbb{P}(213) + \mathbb{P}(312) = (1-b)b$

$$\mathbb{P}(1234) = b^3$$
$$\mathbb{P}(4321) = (1-b)^3$$
$$\mathbb{P}(1243) + \mathbb{P}(1342) + \mathbb{P}(2341) = (1-b)b^2$$
$$\mathbb{P}(1432) + \mathbb{P}(2431) + \mathbb{P}(3421) = (1-b)^2 b$$
$$\mathbb{P}(2134) + \mathbb{P}(3124) + \mathbb{P}(4123) = (1-b)b^2$$
$$\mathbb{P}(3214) + \mathbb{P}(4213) + \mathbb{P}(4312) = (1-b)^2 b$$
$$\mathbb{P}(1324) + \mathbb{P}(1423) + \mathbb{P}(2314) + \mathbb{P}(2413) + \mathbb{P}(3412) = (1-b)b^2$$
$$\mathbb{P}(4231) + \mathbb{P}(3241) + \mathbb{P}(4132) + \mathbb{P}(3142) + \mathbb{P}(2143) = (1-b)^2 b$$

Hyperplanes

Example

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

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

Hyperplanes

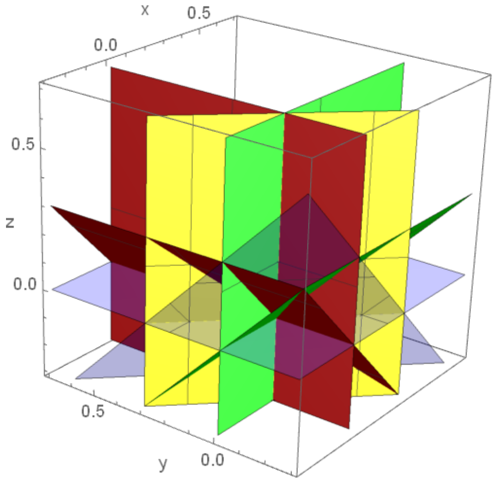
Example

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

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

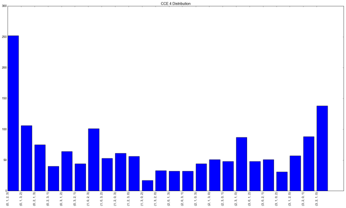
π	$\mathbb{P}(\pi)$	π	$\mathbb{P}(\pi)$
123	b^2	132	$(b-1) - \frac{3}{2}(b-1)^2$
213	$(b-1) - \frac{3}{2}(b-1)^2$	231	$\frac{1}{2}(b-1)^2$
312	$\frac{1}{2}(b-1)^2$	321	$(b-1)^2$

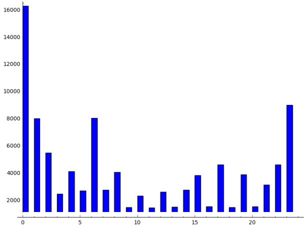
Integration Regions



Dartmouth

Stock Comparison





Equivalence Classes

Theorem (Martinez 2015)

Let $\pi, \tau \in S_n$. The $\mathbb{P}(\pi) = \mathbb{P}(\tau)$ for every probability distribution on the $\{X_i\}$ if and only if π and τ are equivalent under repeated application of the reverse complement operation on bounded cylindrical blocks.

Equivalence Classes

Theorem (Martinez 2015)

Let $\pi, \tau \in S_n$. The $\mathbb{P}(\pi) = \mathbb{P}(\tau)$ for every probability distribution on the $\{X_i\}$ if and only if π and τ are equivalent under repeated application of the reverse complement operation on bounded cylindrical blocks.

{123}, {132, 213}, { 231, 312}, {321}
{1234}, {1243, 2134}, {1324}, {1342, 3124}, {1423, 2314}, {1432, 2143, 3214}, {2341, 3412, 4123}, {2413}, {2431, 4213}, {4231}, {3142}, {3241, 4132}, {3421, 4312}, {4321}
{12345}, {14325}, {21354}, {21453}, {25314}, {41352}, {45312},{52341}, {54321}, {12543, 32145}, {13245, 12435}, {13425, 14235}, {15243, 32415}, {15342, 42315}, {15432, 43215}, {21345, 12354}, {21435, 13254}, {21543, 32154}, {23145, 12534}, {23415, 15234}, {24153, 31524}, {24315, 15324}, {24513, 35124}, {24531, 53124}, {25134, 23514}, {25341, 52314}, {25413, 35214}, {25431, 53214}, {31245, 12453}, {31425, 14253}, {31542, 42153}, {32514, 25143}, {32541, 52143}, {35142, 42513}, {35241, 52413}, {41235, 13452}, {41253, 31452}, {41325, 14352}, {41523, 34152}, {41532, 43152}, {42135, 13542}, {42351, 51342}, {45231, 53412}, {51324, 24351}, {51423, 34251}, {51432, 43251}, {53142, 42531}, {53241, 52431}, {54123, 34521}, {54132, 43521}, {54213, 35421, }, {54231, 53421}, {54312, 45321}, {31254}, {43125, 14532}, {34125, 14523}, {13524, 24135}, {35412, 52134, 45213, 23541}, {51243, 32451} {43512, 45132}, {23451, 45123, 34512, 51234} {21534, 23154, 15423, 34215}

Related Metrics

Definition

To measure how closely the distribution matches the conditions of Martinez we compute

$$g_n(T) = \sum_{\Lambda_i \subset S_n} \sum_{\pi \in \Lambda_i} p_\pi |p_\pi - \mu_i|,$$

or "equivalently",

$$h_n(T) = \sum_{\Lambda_i \subset S_n} \sum_{\pi \in \Lambda_i} p_\pi \log\left(\frac{p_\pi}{\mu_i}\right).$$

Related Metrics

Definition

To measure how closely the distribution matches the conditions of Martinez we compute

$$g_n(T) = \sum_{\Lambda_i \subset S_n} \sum_{\pi \in \Lambda_i} p_\pi |p_\pi - \mu_i|,$$

or "equivalently",

$$h_n(T) = \sum_{\Lambda_i \subset S_n} \sum_{\pi \in \Lambda_i} p_\pi \log\left(\frac{p_\pi}{\mu_i}\right).$$

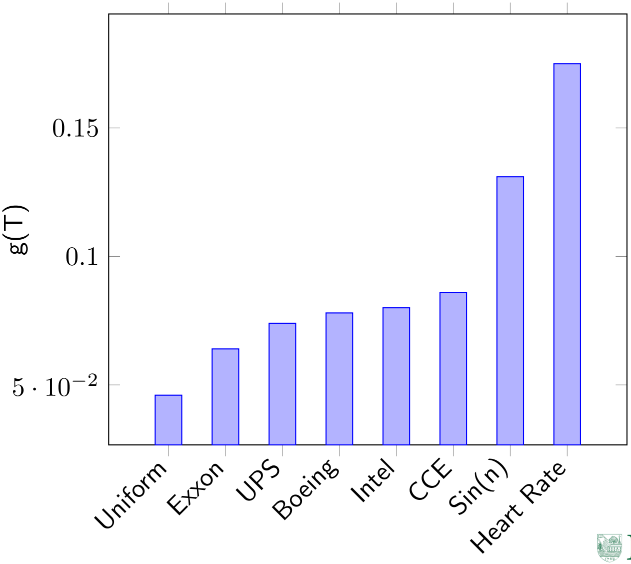
Corollary

If $Z^N = \{Z_i\}_{i=1}^N$ *is a random walk, then*

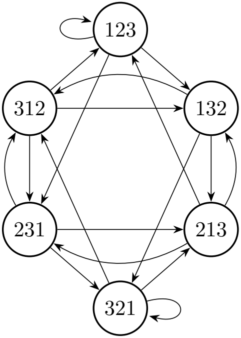
$$\lim_{N\to\infty} g_n(Z^N) = 0 \textit{ and } \lim_{N\to\infty} h_n(Z^N) = 0.$$

uth

Examples


g(T)
0.15
0.1
$5 \cdot 10^{-2}$
Uniform
Exxon
UPS
Boeing
Intel
CCE
Sin(n)
Heart Rate

Beyond Distributions



(a) The permutation graph for $n = 3$ whose edges, $\pi \to \tau$, are weighted with probability $P_X(\pi \to \tau)$.

Uniform Walks

Lemma

Let $\pi, \tau \in S_n$. *Then,*

$$\mathbb{P}(\tau \to \pi) = \frac{\mathbb{P}(\tau \wedge \pi)}{\mathbb{P}(\tau)}$$

Uniform Walks

Lemma

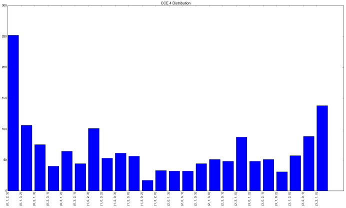
Let $\pi, \tau \in S_n$. Then,

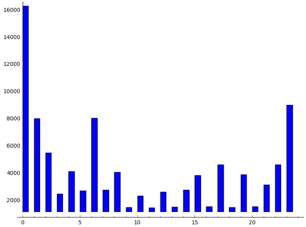
$$\mathbb{P}(\tau \to \pi) = \frac{\mathbb{P}(\tau \wedge \pi)}{\mathbb{P}(\tau)}$$

Lemma

Let $\{Z_i\}$ be a random walk as above. Then, the transition probability between permutations are not the uniform distribution.

CCE Random Walk





Further Reading

- M. Riedl, A. Müller, and N. Wessel: *Practical considerations of permutation entropy*, Eur. Phys. J. Special Topics 222, 249262 (2013).
- 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.
- L. Zunino, M. Zanin, B. M. Tabake, D. G. Prez, O. A. Rosso: *Forbidden patterns, permutation entropy and stock market inefficiency*, Physica A 388 (2009) 2854-2864.

Code (Try it yourself...)

https://math.dartmouth.edu/~ddeford/time_series.html

That's all...

Thank You!