Markov Chains:
using MC and HMM to describe and explore time series data

Stephen J. Merrill
Department of MSCS
Program in Computational Sciences
Marquette University

What to take home from this talk

- Markov chains (MC) and the related idea of Hidden Markov Models (HMM) are feasible ways to

1. Describe and analyze the behavior of both categorical and numeric time series
2. Provide simulations which preserve the correlation structure of the original series
3. Describe changes (bifurcations) in the original series
4. Study the impact of noise

Dynamic "Modeling"

Constructing a mathematical model for a dynamic process can follow several paths

- **First Principle Models** – where the "physics" of the system dictates the form of the equations.

Example : equation of motion of a spring based on Hooke's Law (restoring force proportional to amount of deformation) and F=ma or a birth-and-death process.

- **Empirical Models** – where data or observations dictates the model within a class of possible models.

Examples : classical Box-Jenkins time series (ARIMA) models or models based on regression.

Deterministic / Stochastic

- Another decision is to whether the process is **essentially deterministic** (and stochastic effects can be ignored) or **one in which stochastic effects are embedded in the dynamics**.
- The difference is usually in the assumption that $x_{n+1} = f(x_n)$ and $y_n = x_n + \varepsilon$ where the y_n are measured with error (but that error does not affect the dynamics) or $x_{n+1} = f(x_n + \varepsilon_n)$ where it does.

Another Possibility

Stochastic models can also be described where there is no such function f. In these cases, <u>the probability of the next value is given by a probability density function depending on</u> x_n .

- Examples: The extreme case is where the sequence is a run of random numbers from the same distribution. Here, the distribution of the next value does not depend on the previous value.
- Markov chains, where the distributions are specified depending on x_n – in the empirical case by the history of the data values.

This Talk

- Brief description of Markov chains
- Estimating transition probabilities from a time series
- Applications of this technique
- Heart rate variability
- Cardiac imaging registration
- Study of chaotic systems
- Hidden Markov Models (HMM)
- Summary and References

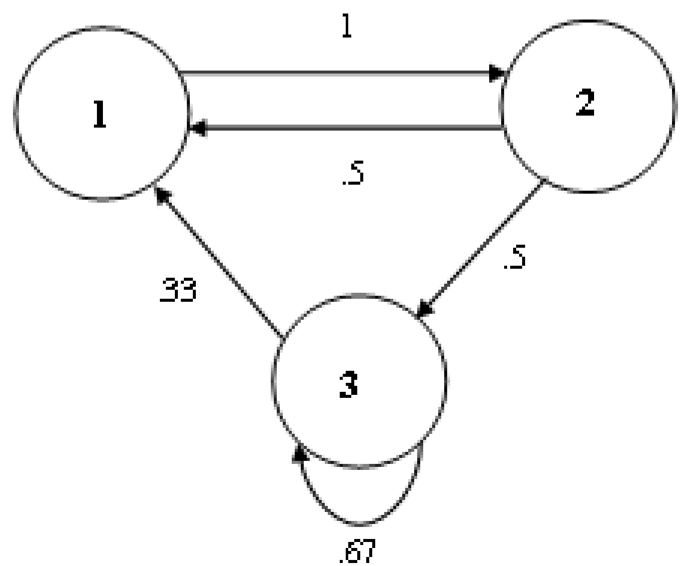
Common Markov Chain Model Assumptions

- **Discrete time – discrete state**
- **Order 1** (only the past state value determines the distribution) – this is the Markov assumption (previous history not relevant).
- If system is continuous state, **need to define what the states are** (think histogram).
- Time homogeneous chains. The model is specified **by determining the transition probabilities** (by first principles or through data) that do not change over time.

The Transition Matrix $\mathbf{P}$

- Current state -> row.
- Possible next state -> columns
- Each row represents a pdf given a current value in that state.

$$\mathbf{P} = \begin{pmatrix} 0 & 1 & 0 \\ .5 & 0 & .5 \\ .33 & 0 & .67 \end{pmatrix}$$



Basic Assumptions – Empirical Chain

- Considering a time series, $\{x_n\}_{n=1}^{N}$, assume that the values are correlated – not independent as in a sequence of random numbers.
- The Markov assumption holds – only the current value is needed to determine the probability distribution for the next (order 1).
- There is sufficient data to approximate these probability distributions based on the number of states.

Approach to estimating P

- Discretize the state space into n states. The choice of n is limited by the data available.
- Estimate the1-step transition probabilities by computing the relative frequencies.

$$p_{ij} = \left\{ \frac{\text{number of times that } x_{k+1} = j}{\text{total number of times that } x_k = i} \right\}$$

- Use the nxn matrix $P = \left(p_{ij} \right)$
 for simulation and analysis
- On the next slide, an AR(1) process is used to generate a time series d to illustrate.

$$x_{n+1} = ax_n + \varepsilon$$

Example of Empirical Transition Matrix

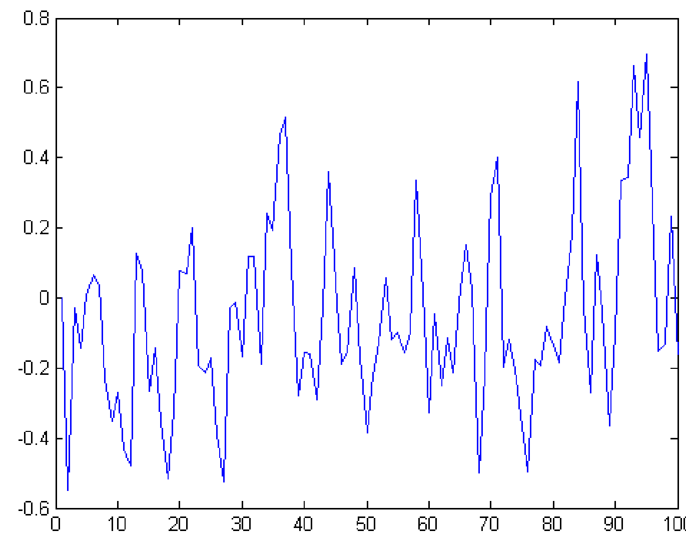
```
>> P=transi(d,5)
```

P =

y next state

x

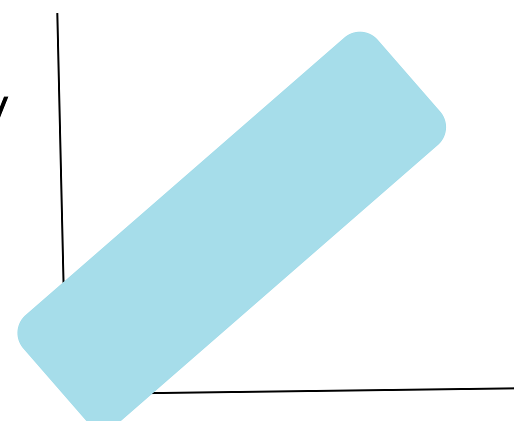
0.3333	0.3333	0.3333	0	0
0.1538	0.5128	0.2051	0.1282	0
0.1379	0.3793	0.3448	0.0690	0.0690
0	0.4000	0.3000	0.2000	0.1000
0	0	0.3333	0.1667	0.5000



n

Note that the pattern of nonzero values in the transition matrix suggests that the form of the relationship f. This would also appear in a lag 1 map.

y



x

Generating a time series from a transition matrix

- Using a Matlab m-file, starting with 1, 10 random numbers based on the matrix P can be generated.
- `>> x=generate(1,P,10)`

```
x =    1     2     3     3     3     1     2     1     2     3
```

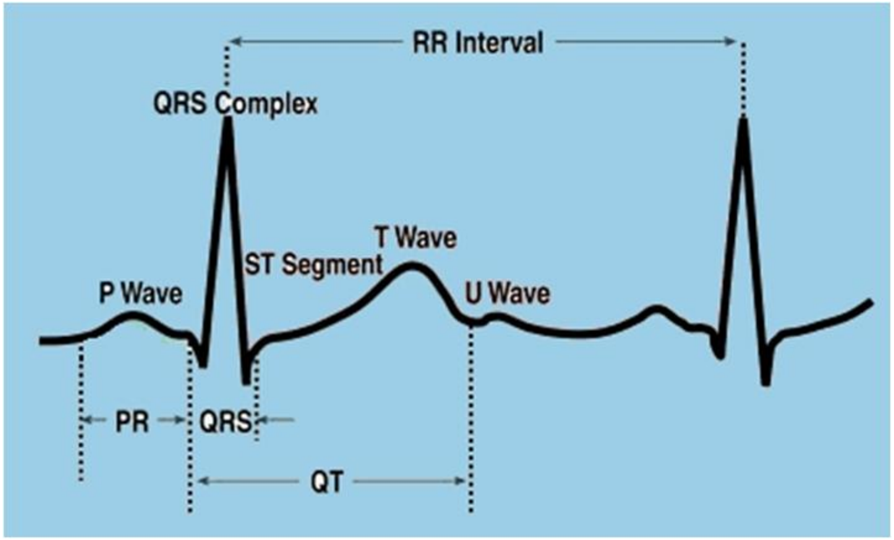
This sequence will have the same correlation structure as the original sequence that may have generated P. This is a very handy way of starting with only one series, and producing others for testing and statistical analysis.

Example – Heart Rate Variability

- Heart Rate Variability (HRV) describes the beat-to-beat variation in the time intervals between beats (the R-R intervals) as seen on ECG. It is described by many different indices.
- The variability is due to several different control mechanisms in the systems.
- Operation of the controls are affected by drugs (specifically here, anesthesia)

Motivation and Background

- Pediatric patients undergoing surgery
- The goal was to design a real-time monitor to anticipate sudden cardiac arrest . Data had been collected on several patients and standard indices did not behave well as measures of HRV in several patients



The HRV data

2.5 year old girl

7.5 year old boy -- early phase with halothane late phase with the addition of atropine

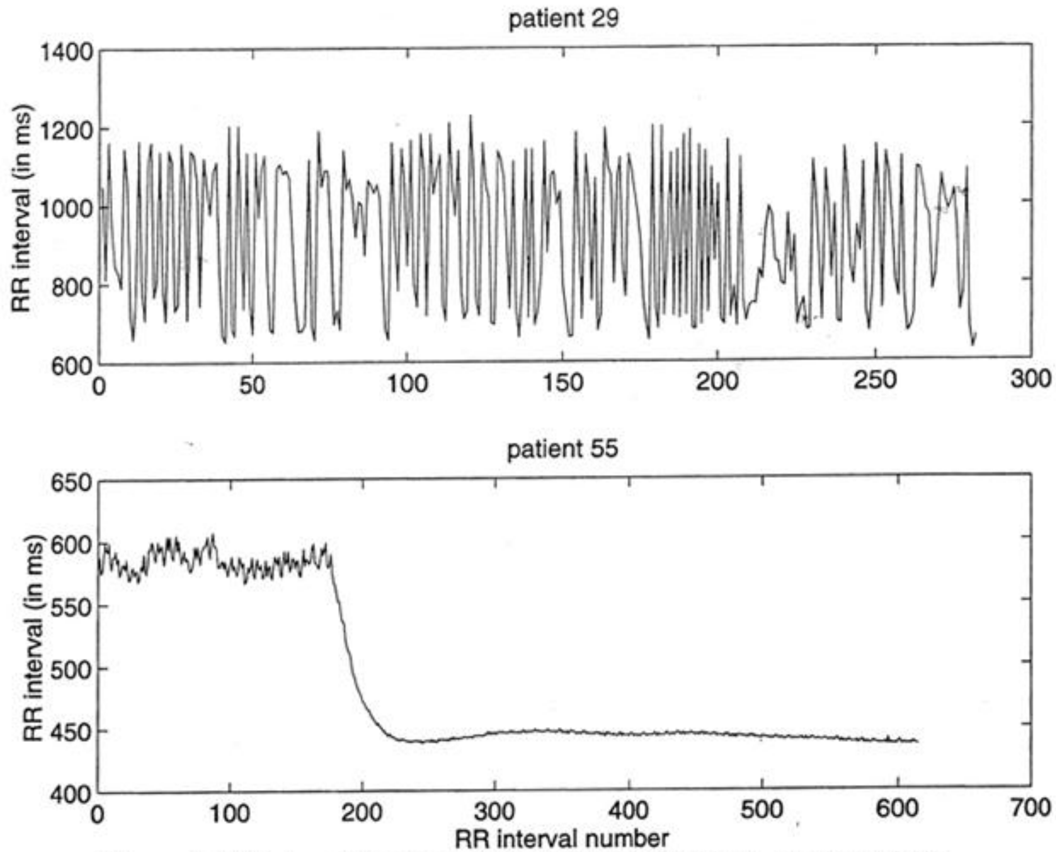
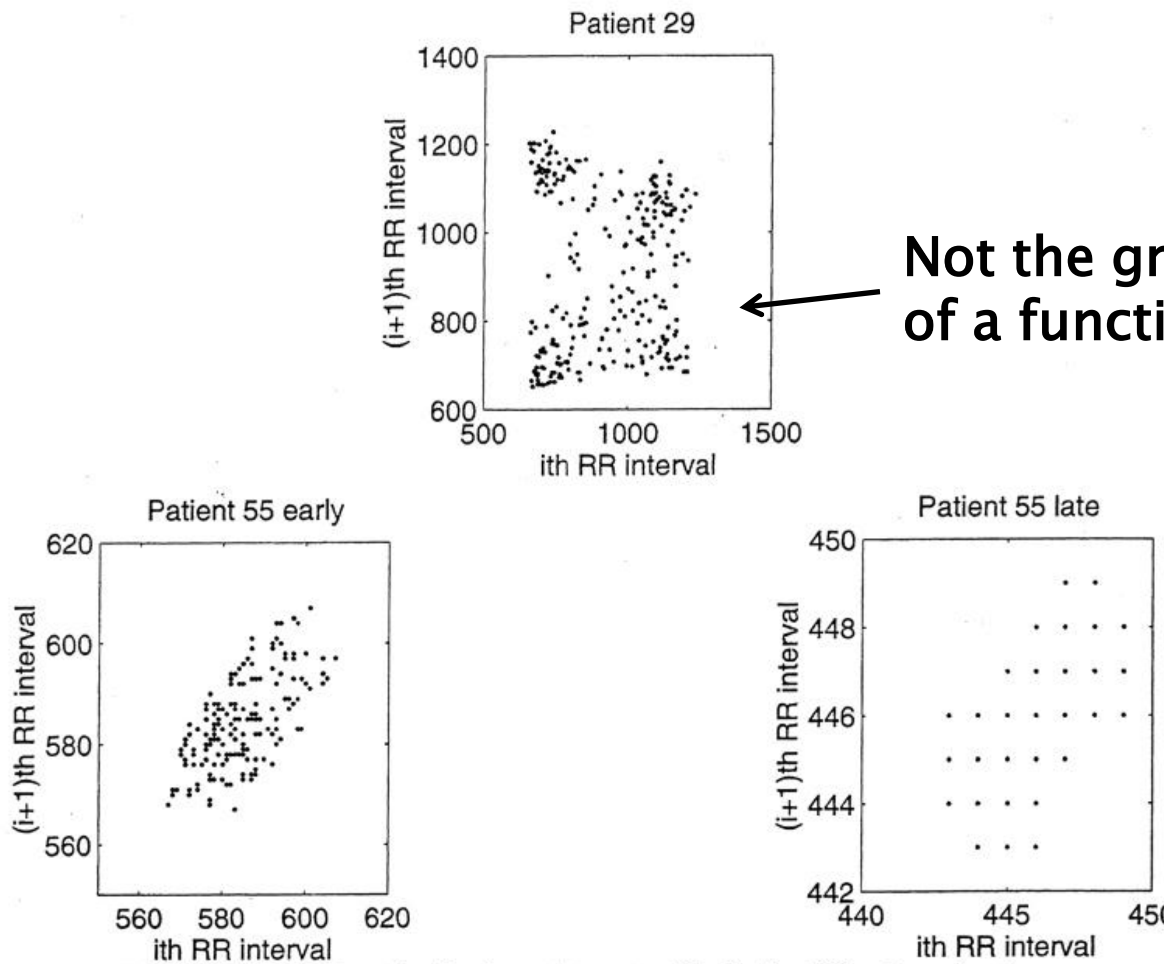


Figure 1 RR interval data from Patient 29 and Patient 55. In the analysis, the second set will be divided into an early and late phase.

Lag 1 Maps – HRV data



Not the graph of a function

Eigenvalues of HRV MC's

- All eigenvalues of a Markov chain transition matrix lie in (or on) the unit circle. Uniqueness of a modulus 1 eigenvalue means the presence of a limit distribution

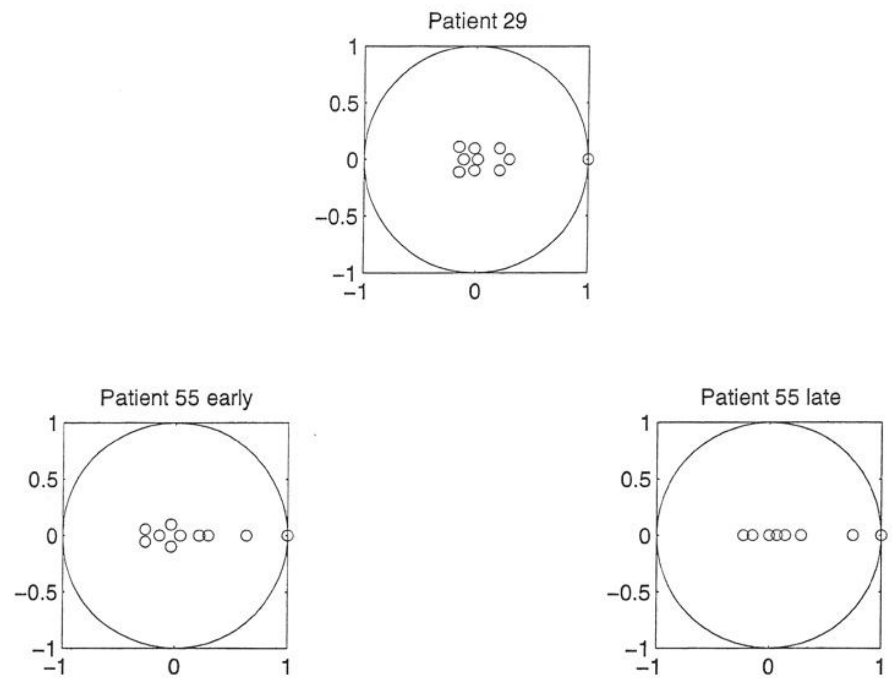


Figure 5 Eigenvalues for each of the three transition matrices are shown in relation to the unit circle. Besides the nearness of the "non-1" eigenvalues to the unit circle, notice the differences in the number of complex eigenvalues.

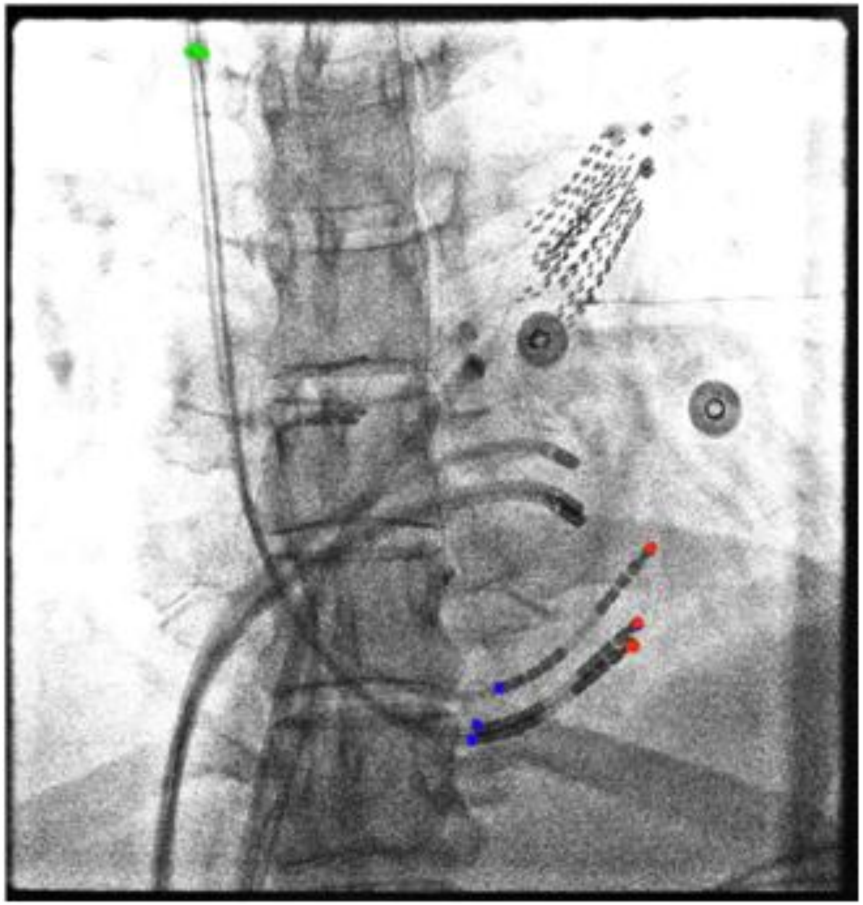
Advantages/disadvantages of the Markov chain model

- Qualitative properties of the chain (e.g. eigenvalues or limit distribution) can be used to characterize the data set or identify changes in the data set.
- Simulation of the chain can result in an unlimited number of sample paths with the same dynamic behavior of the original set.
- Transition matrix depends on the definition of the "bins" and the number of them.
- N states requires estimating N^2 probabilities

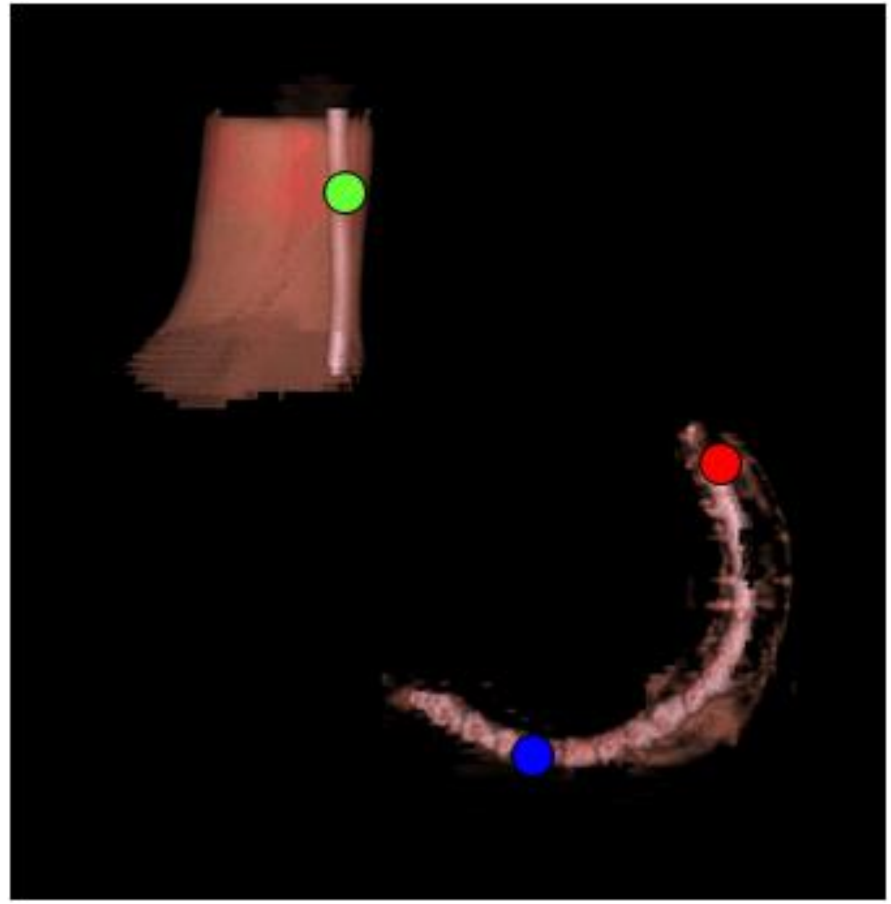
Cardiac Image Registration
Shivani Ratnakumar

- Register 2-d (real-time fluoroscope) to 3-d (static 3-d CT) image to guide ablation catheter in treatment of atrial fribrillation.
- The selection of corresponding fiducial points is difficult in the moving image (even with gating for cardiac cycle and breathing)
- The chaotic movement of points in the heart make a model problematic. The thought was to create an empirical Markov chain to describe this movement.

The Registration Problem



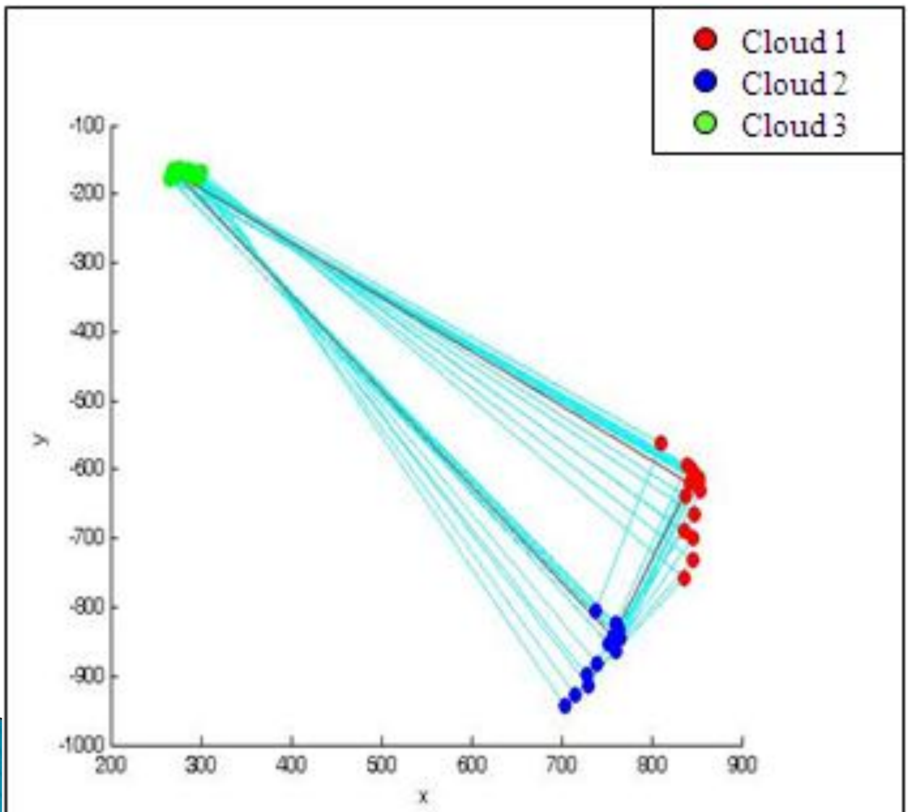
Three layered ECG gated fluoro frames



Segmented CT image that will be registered

Concentrating on the dynamics

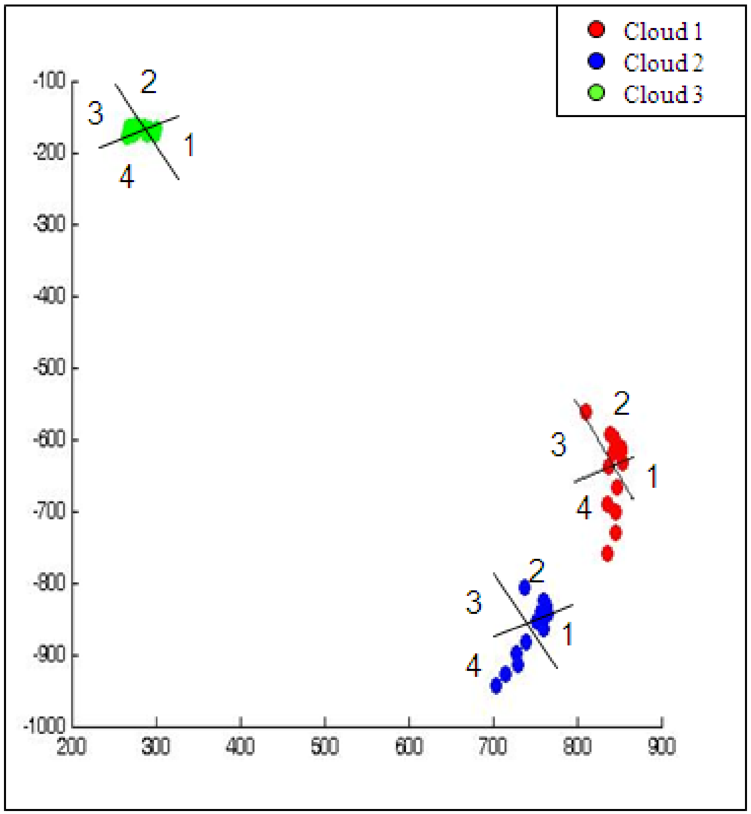
ECG gated fluoro data -- triples of points, each forming a “cloud”
2d coordinate locations of three points were recorded across fluoro sequences



$$
\begin{array}{ccc}
1 & 2 & 3 \\
\end{array}
$$

$$
\begin{bmatrix}
(852,-618) & (766,-844) & (266,-177) \\
(810,-562) & (738,-806) & (268,-164) \\
\vdots & \vdots & \vdots \\
(848,-608) & (760,-836) & (273,-174)
\end{bmatrix}
$$

Discretize the state space



$$\begin{array}{c} \begin{matrix} \color{red}{1} & \color{blue}{2} & \color{green}{3} \end{matrix} \\ \begin{bmatrix} 2 & 2 & 4 \\ 2 & 2 & 3 \\ \vdots & \vdots & \vdots \\ 2 & 2 & 4 \end{bmatrix} \end{array}$$

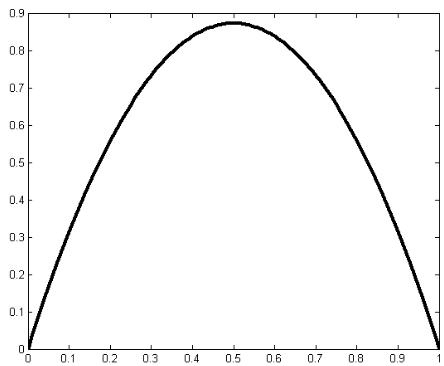
Sequence of coordinate points triplets can be written as a sequence of state triplets

Building a Transition Matrix

- **For each row**, there is a probability that a "1" in cloud 1 is associated with a "1" in cloud 2. Similarly for cloud 2 to cloud 3 and cloud3 to cloud 1 in the next row (next time). Compute these probability from the data (~30 rows).
- From these matrices, **compute cloud 1 to cloud 1 one step transition probabilities** (details omitted).
- **This 1 -> 1 matrix is a description of the motion of that area of the heart.**
- This was used (through the limit distribution of that matrix) to find a well-defined fiducial point.

Using Markov Chains to study chaotic systems

- Example: bifurcation in the discrete logistic equation:



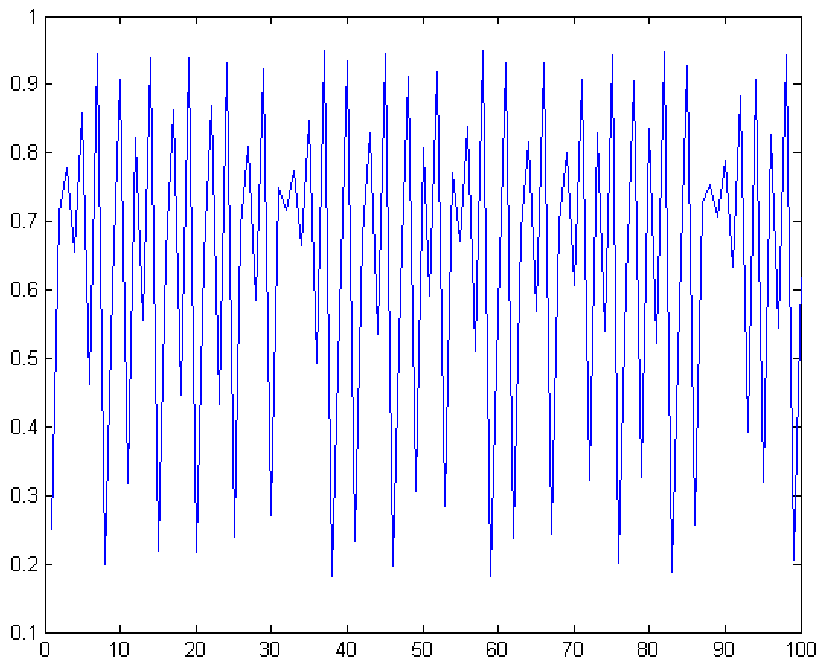
$$x_{n+1} = \mu x_n (1 - x_n) \quad 0 \le \mu \le 4$$

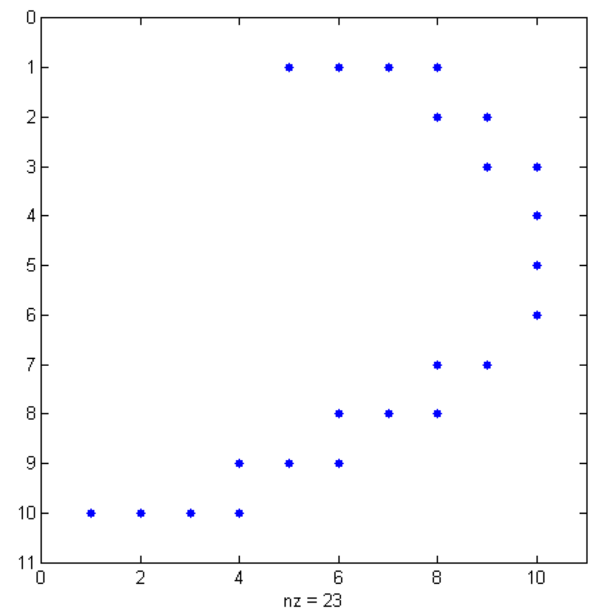
- Summary of behavior: $x = 0$ is always an equilibrium, stable if $\mu < 1$.

  $x = 1 - \frac{1}{\mu}$ is an equilibrium that is stable when $1 < \mu < 3$

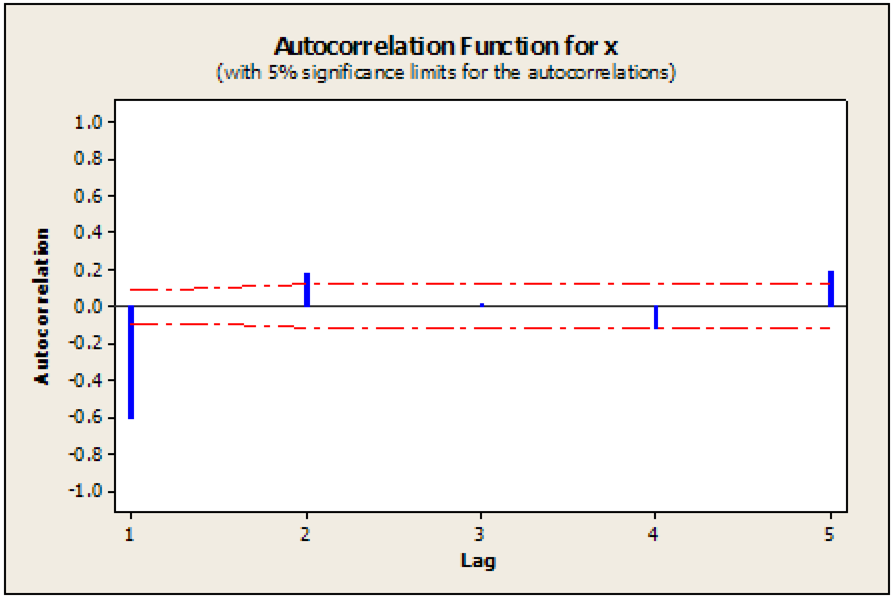
-- then the fun starts.

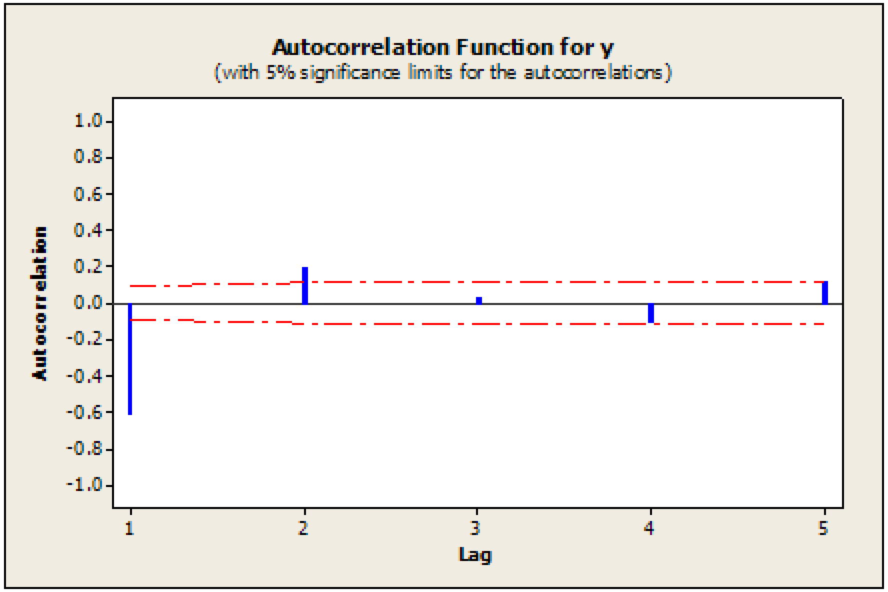
Transition matrix





Nonzero entries in the transition matrix with 10 states $\mu = 3.8$


Autocorrelation Function for x
(with 5% significance limits for the autocorrelations)
Autocorrelation
1.0
0.8
0.6
0.4
0.2
0.0
-0.2
-0.4
-0.6
-0.8
-1.0
1
2
3
4
5
Lag

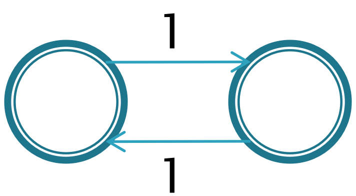

Autocorrelation Function for y
(with 5% significance limits for the autocorrelations)
Autocorrelation
1.0
0.8
0.6
0.4
0.2
0.0
-0.2
-0.4
-0.6
-0.8
-1.0
1
2
3
4
5
Lag

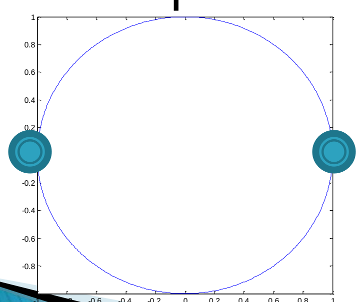
Bifurcation in Markov Chains

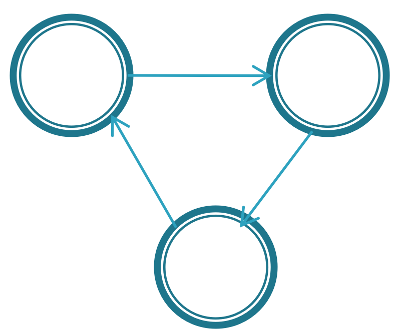
- One can use the empirical chain to detect bifurcations. In this case, the bifurcation at $\mu = 3$ to a period 2 point. This is done through the eigenvalues of the transition matrix.
- Facts: 1 is always an eigenvalue and others are on or inside the unit circle. If 1 is the only eigenvalue of magnitude 1, there is a unique limit distribution.
- Loss of the unique limit distribution and establishment of a new one is the indication of a bifurcation in the system.
- As a parameter changes, look for **eigenvalue(s)**
- **approaching the unit circle**.

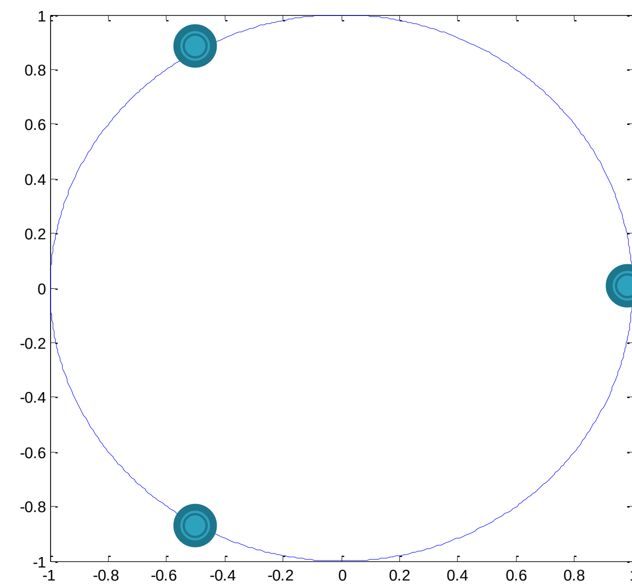
Meaning of the eigenvalues

- The eigenvalues describe the approach to a limit distribution
- The eigenvalues also describe the modes that are available – these depend on the nature of the digraph of transitions.









Bifurcation to a period 2 point

- Transition matrix computed with 10 states at $\mu = 2.9, 3.0, \text{and } 3.05$
- At 2.9, {–.81, .81, 1, lots of 0's}
- At 3.0, {–.92,.917, 1, lots of 0's}
- At 3.05, {–1,1, lots of 0's}
- From the above, it appears that a bifurcation took place just before 3.05. Moreover, because of the –1, it is a period 2 point.

The next bifurcation

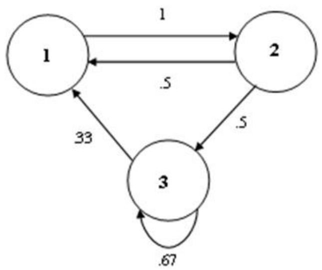
- Transition matrix computed with 10 states at $\mu = 3.5,\ 3.503,\ \text{and}\ 3.6$ and eigenvalues computed
- At 3.5, {−1, 1, lots of 0’s}
- At 3.503, {−1,1, 0.0000 + 0.1010i, 0.0000 – 0.1010i, lots of 0’s}
- At 3.6, {−1,1, i, −i, lots of 0’s}

From the above, it appears that a bifurcation took place just before 3.6. Moreover, because of the i, it is a period 4 point.

Hidden Markov Models (HMM)

- The idea here is that there is data "**emissions**", the states observed in the system. It is expected --- usually because other models did not fit – that the observed values depend on the state of a non-observed process.
- In the HMM, the non-observed process is a MC.
- To specify a HMM, one has a **transition matrix** for the MC and a **matrix of emission probabilities – the probability of a particular emission given the state of the MC**

Example



- Suppose the MC in slide 8. There are 3 (hidden) states and a transition matrix P.
- Suppose the emissions are 1 or 2 with emission probabilities

$$\text{emis} = \begin{bmatrix} .5 & .5 \\ 1 & 0 \\ 0 & 1 \end{bmatrix}$$

The MATLAB function hmmgenerate will generate a sequence (of length len) of emissions and associated states.

- Note that this is our function **generate** if emis is the identity matrix.

HMM -- questions

- From a sequence of emissions and states, estimate the two matrices. This is done by hmmestimate – similar idea to estimating a transition matrix (twice)
- Given a sequence of emissions, estimate the matrices – here you are guessing as to the number of states in the MC. This is done by hmmtrain.
- Popular to answer: is a given sequence likely to have been generated by a given process?
- Further examples on HMM on the CD

Summary

- Markov chains can be employed without specifically knowing the (functional) nature of the dynamics.
- Having a Markov chain model enables one to simulated the process – generating new time series with the same properties as the original.
- The transition matrices and the associated limit distribution and eigenvalues contain useful information on the process.
- HMM extends this idea to cases when the observed state is probabilistically determined by a Markov chain.

References

- Merrill, S.J. (2011) Markov chains for identifying nonlinear dynamics, in Nonlinear Dynamical Systems Analysis for the Behavioral Sciences Using Real Data, S. Guastello and R. Gregson (eds), CRC Press, 401–423.
- Merrill, S.J. & Cochran, J.R. (1997) Markov chain methods in the analysis of heart rate variability, *Fields Inst Comm* **11**:241–252.
- Nicolis, G. & Nicolis, C. (1990) Chaotic dynamics, Markovian coarse-graining and information, *Physica A* **163**:215–231.
- Pincus, S.M. (1991) Approximate entropy as a measure of system complexity, *Proc Natl Acad Sci USA* **88**:2297–2301.
- Pincus, S.M. (1992) Approximating Markov Chains, *Proc Natl Acad Sci USA* **89**:4432–4436.
- Ratnakumar, S. (2008) Markov chain modeling of ECG gates live left atrial fluoroscopy variability to establish a well-defined basis for rigid registration to a 3D CT image, *Ph. D. Dissertation, Marquette University.*

On the CD – Matlab files and examples

- **generate.m**

This function takes an initial state and transition matrix P of a Markov Chain and

produces a sample path of length n.

- **logistic.m**

A function to generate time series from the discrete logistic equation.

- **transi.m**

This m-file takes a time series d of floating point numbers and creates an empirical mxm Markov chain transition matrix.

- **ploteig.m**

plot of eigenvalues of the transition matrix in relation to the unit circle.

- **HMM examples** using the Statistics Toolbox.