Using Markov Chains for Modeling, Simulation, and Analysis

Stephen Merrill
Department of MSCS
Marquette University
Milwaukee, WI

The idea of the workshop

- **Using Markov chains as general tools** for building and analyzing models and time series is described.
- **What is new** here (and not published) is the application of these tools to **analyze the dynamical behavior of simple systems** – including deducing the structure of the system and possible bifurcation behavior.
- These systems are richer than the classic chaotic systems in that all 1-d dynamical systems can be described by these methods, but not all Markov chains can be seen as a 1-d system. In addition, the system is allowed to contain noise.

What you might learn

- Some MATLAB. We will depend heavily on the standard MATLAB features – supplementing with M-files provided on your CD.
- Some of the features of Markov chain models – simulating a given model and estimating a model from a time series.
- Analyzing chaotic time series that may or may not contain noise. Also, the role of noise can be evaluated in some cases.
- Hidden Markov Models. This workshop uses tools in the Statistics toolbox of MATLAB.

This Workshop

- Introduction to MATLAB
- Models of dynamics
- Examples of Markov chains (MC)
- Eigenvalues, stationary vectors, and limit distributions
- Estimating transition probabilities from a time series & simulating a given MC
- Hidden Markov Models (HMM)
- Applications of this technique
- Summary and References

MATLAB

- MATLAB is the primary product of the Mathworks (http://www.mathworks.com) , in business since 1984.
- MATLAB consists of a base and lots of toolboxes, some application-specific, some third party, as well as a very large collection of user-supplied software in the form of "M-files." These files can be macros (bundled list of commands) or functions (where parameters are passed and returned).

Basic MATLAB features

- Interactive interface – interpretive
- Built-in editor for M-file building and testing
- Rich collection of built-in mathematical and statistical functions
- Algorithms based on up-to-date code
- Good collection of graphics options/editing
- Versions available for many platforms
- M-files are device-independent
- MATLAB can take advantages of multicore and parallel architectures

MATLAB syntax

- Case sensitive – assumes most quantities are matrices
- Most common errors are “shape” issues
- Multiplications are assumed to be matrix multiplications – term-by-term multiplications require a “.” before the operator. Example 1, see the notes.
- Indices start at 1 (not zero)
- Good help facility and docs online.

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 a surface determined through regression.

Example 1

- Flip a fair coin. We would expect that

  $$P(H) = P(T) = \tfrac{1}{2}$$

  and that the previous result would not affect the next. We can generate a sequence of H and T or 1 and 2 by using a discrete uniform random number generator (on MATLAB, randi):

```
1   2   1   2   2   1   2   2   2   2
```

Example 1 Continued

- On the other hand, if one received a sequence of numbers, a time series of 1's and 2's without full knowledge about how it was generated, then one might use what you have

  `1   2   1   2   2   1   2   2   2   2`

  to suggest that the system that produced the 10 numbers seems to generate a "1" 30 % of the time and a "2" 70% of the time. Using this would be an empirical model.

Deterministic / Stochastic

- **Another decision in modeling is 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 $x_{n+1} = f(x_n)$ where the x_n's are measured with error but that **error does not affect the dynamics**

    (the measureable quantity y satisfies $y=x+\varepsilon$)

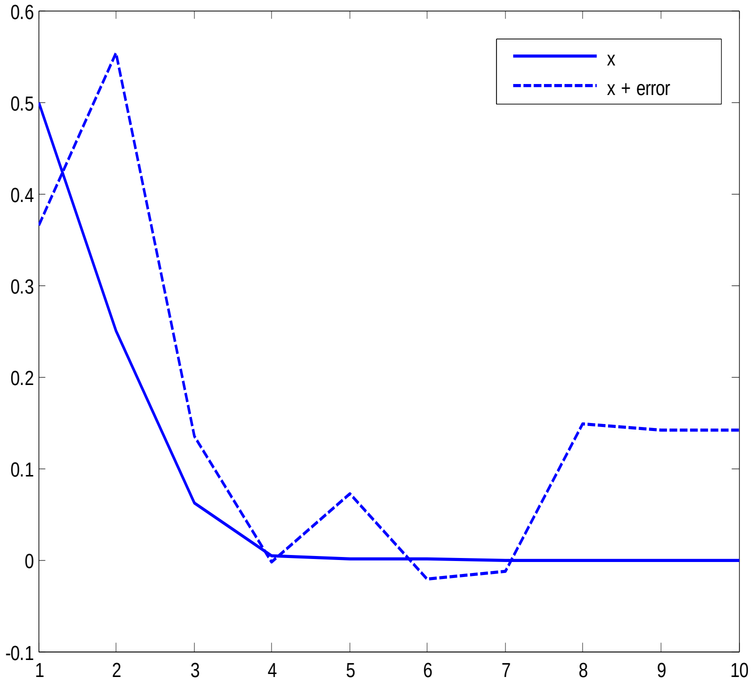
    or $x_{n+1} = f(x_n + \varepsilon_n)$ or $f(x_n) + \varepsilon_n$

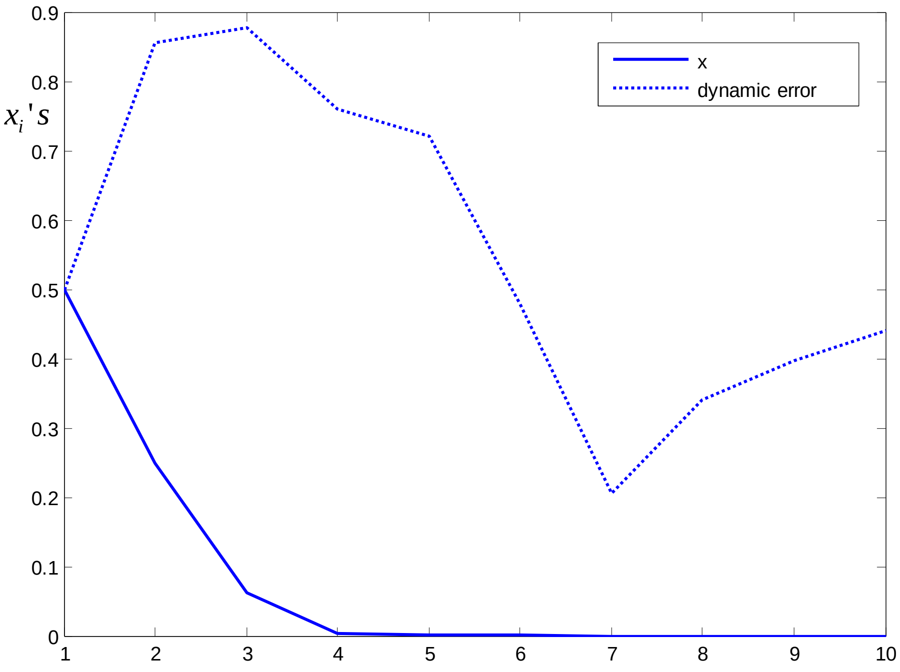
    where the error does affect the dynamics (the sequence of x's).

Additive Error vs Dynamic Error – Example 2

$$x_{i+1} = x_i^2 \, , \, y_i = x_i + \varepsilon_i$$

$$x_{i+1} = x_i^2 + \varepsilon_i$$

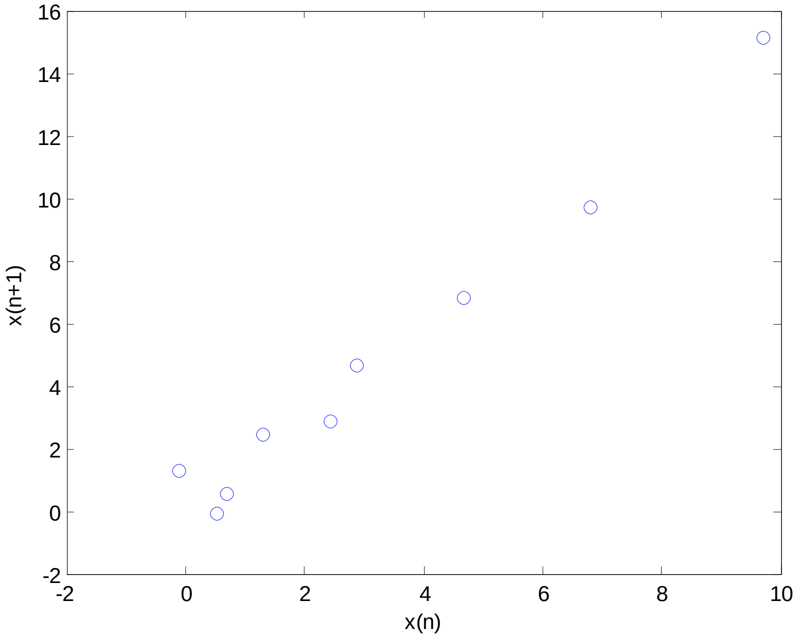




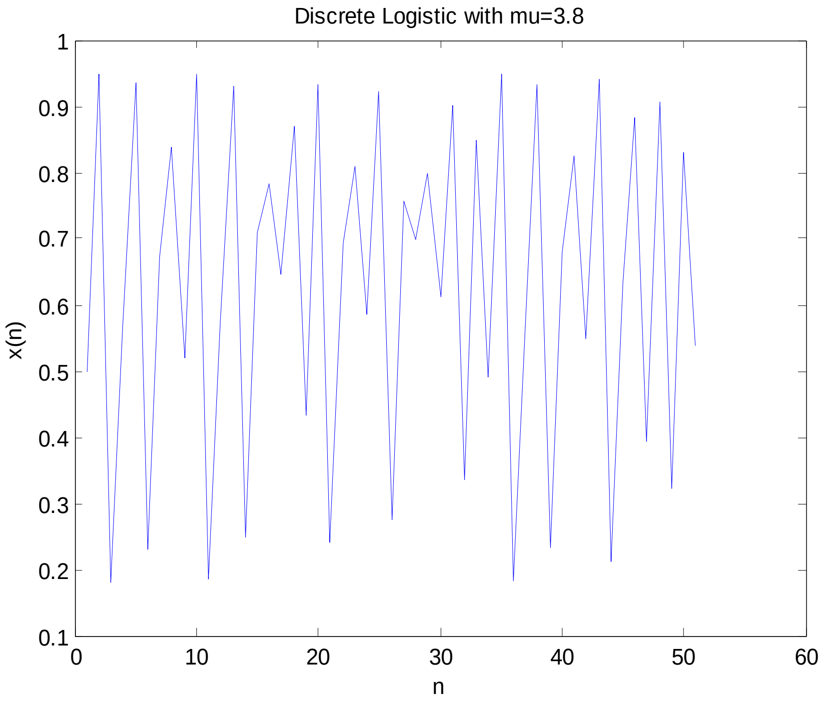
Lag one maps – Example 3

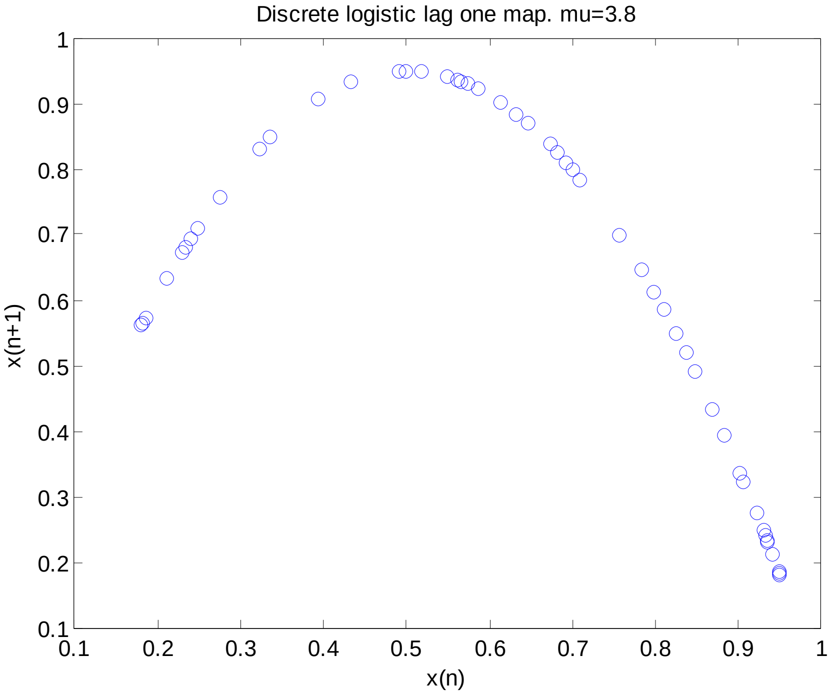
- A lag 1 map is a scatterplot of x(n+1) vs x(n).
- If there is a functional relationship, this plot displays it (and any noise present).
- This example shows that *x(n+1)=mx(n)+b* for some *m* and *b*.

- See code in the notes



In 1-d dynamics, the lag one map is very useful


Discrete Logistic with mu=3.8


Discrete logistic lag one map. mu=3.8

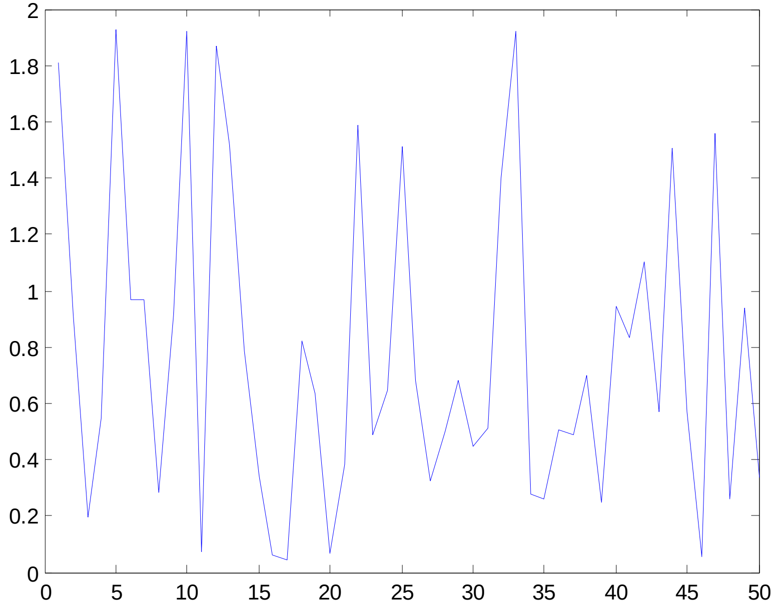
Another Possibility

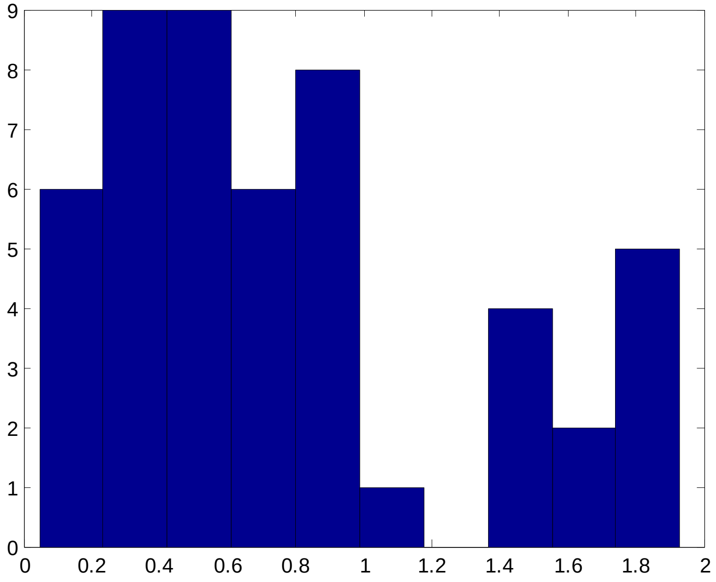
- **Stochastic models can also be described where there is no such function *f*.** For instance, the probability of the next value may be given by a probability density function (pdf) depending on the current or some other variable.
  - Examples: The extreme case is where the sequence is a run of **random numbers** from the same distribution. Here, the distribution does not depend on the previous value or values.

    Markov chains, where the distributions are specified – in the empirical case by the immediate past history of the data values.

Example 4– random numbers

- Suppose a coin is flipped. If a head, a uniformly distributed random number on [0,1] is produced. If a tail, a uniform random number on [0,2] is produced.





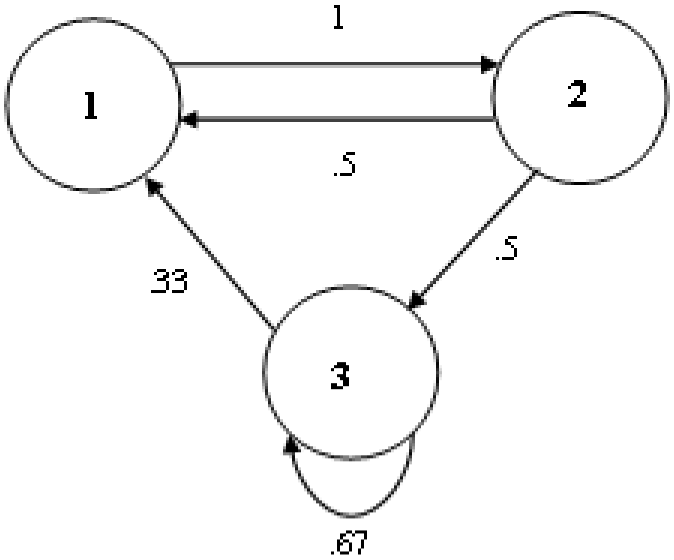
Markov Chain Model Assumptions

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

The Transition Matrix $\mathbf{P}$

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

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



Some linear algebra and MC facts

- An nxn (square) matrix has n **eigenvalues** (counting multiplicity)
- With the transition matrix of a MC, 1 is always an eigenvalue and all eigenvalues are inside or on the unit circle in the complex plane.
- If there is only one eigenvalue on the unit circle, there is a limit distribution.
- This is the left eigenvector of 1.

MATLAB eigenvectors/eigenvalues

```
>> [V,D]=eig(P')

V =
    0.4934    0.1694    0.4825
   -0.8101    0.6070    0.4825
    0.3167   -0.7764    0.7310

D =
   -0.6091         0         0
         0    0.2791         0
         0         0    1.0000
```

```
>> s=sum(V(1:3,3)) ;
  (s = 1.6960)
>> V(1:3,3)/s
ans =
        0.2845
        0.2845
        0.4310
```

This is the **limit distribution** or the **steady-state vector**

Generating a time series from a transition matrix

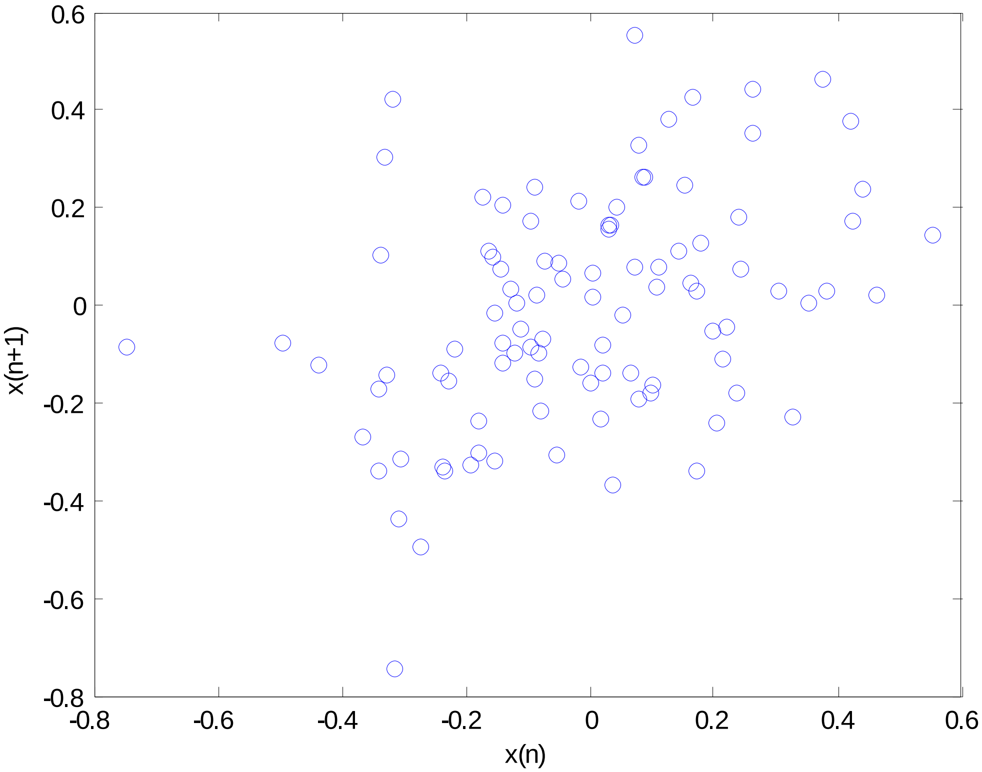
- Your CD contains a number of “M-files.” In this workshop, they will be indicated with **bold** when used.
- Using a Matlab m-file, starting with 1, 10 random numbers based on the matrix P can be generated.
- >> x=**generate**(1,P,10)

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

Each time this is executed, a different sequence will result (all starting at 1).

AR(1) time series – lag 1 map

$$x_{n+1} = .5x_n + \varepsilon_n$$

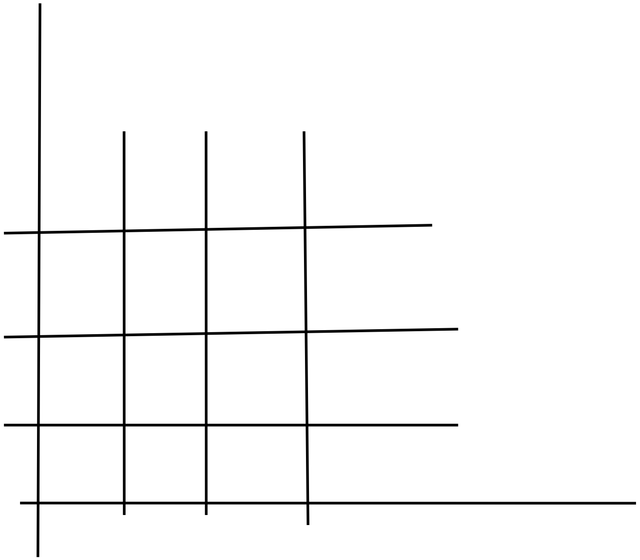


Building a transition matrix from data

- To create a transition matrix from a time series, one needs to first determine the number of states. As in making a histogram, the number of classes is important.
- Within each class (row of the transition matrix) one needs to determine what class (column) the next value falls in. Keep a count.
- Divide each row by the number of counts in that row. If the number of counts is positive.

The basic idea

- With a lag 1 map, divide the horizontal (x(n)) into a number of classes. Divide the vertical axis at the same values.
- Count the number of data points in each square
- For each column, count the fraction of the counts in that column



Building an Empirical Transition Matrix

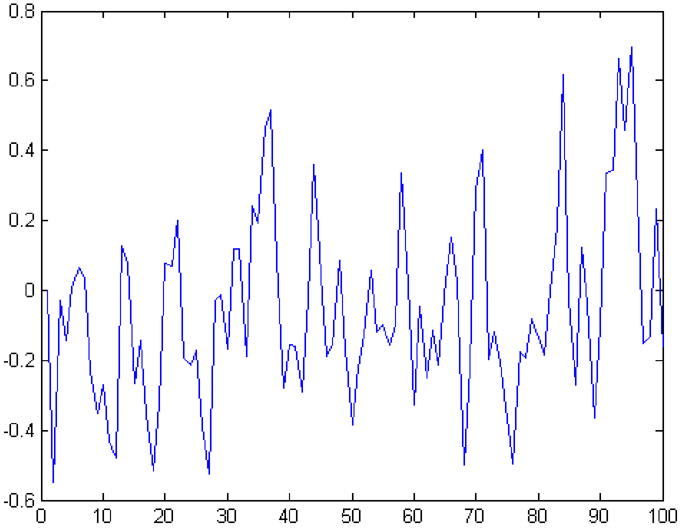
```
>> Dt=transi(d,5)
```

y

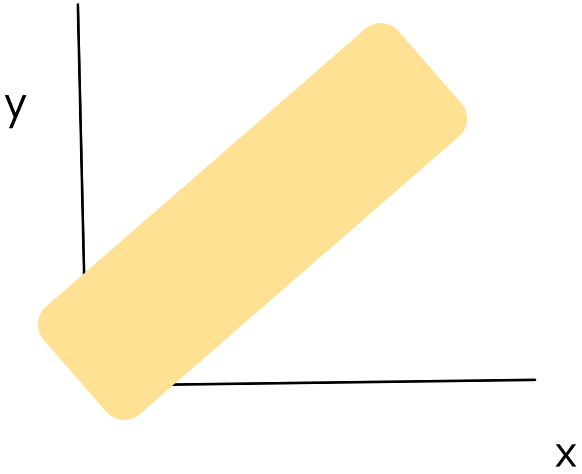
x

```
Dt =

    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
```



Note that the pattern of nonzero values in the transition matrix suggests that the form of the relationship f



Regression on each row (“lag regression”)

- One can sometimes determine a potential relationship between current and next values if given a transition matrix – this assumes that the next value is a function of the previous.
- The usual approach is for each row, compute the mean next value – then do a regression on these values. (For each row, look at $(x, \bar{x})$)

Hidden Markov Models (HMM)

- The idea here is that there is data "**emissions**" 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 4

- Suppose the MC in slide 18. There are 3 (hidden) states and a transition matrix P.
- Suppose the emissions are 1 or 2 with emission probabilities
- emis=[.5,.5;1,0;0,1]
- 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.
- 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.

Applications of these methods

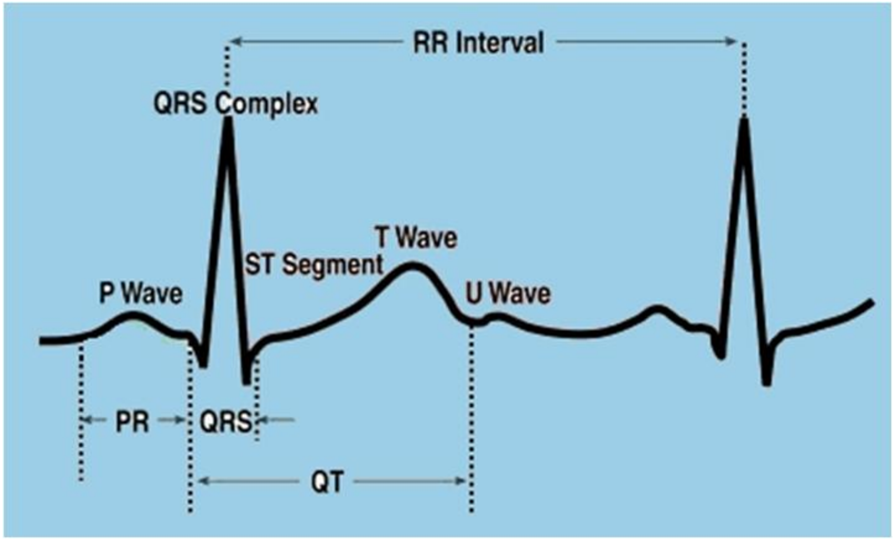
- Heart rate variability
- Heart image registration
- Bifurcation in 1-d dynamics
- Spatial variation in the microarray

Heart Rate Variability

- Heart Rate Variability (HRV) describes the beat-to-beat variation in the time interval between beats 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 -- <u>early phase</u> with halothane <u>late phase</u> with the addition of atropine

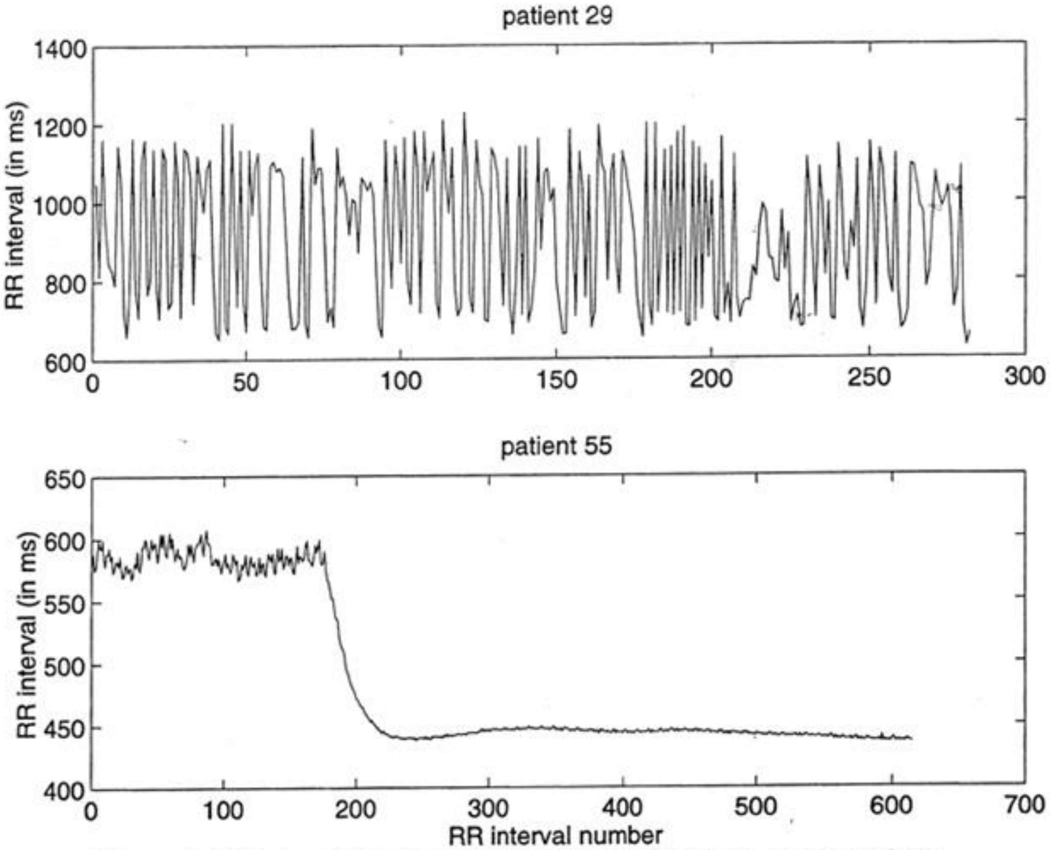


Figure 1 RR interval data from Patient 29 and Patient 55. In the analysis, the second set will be divided into an early and late phase.

Lag 1 Maps – HRV data

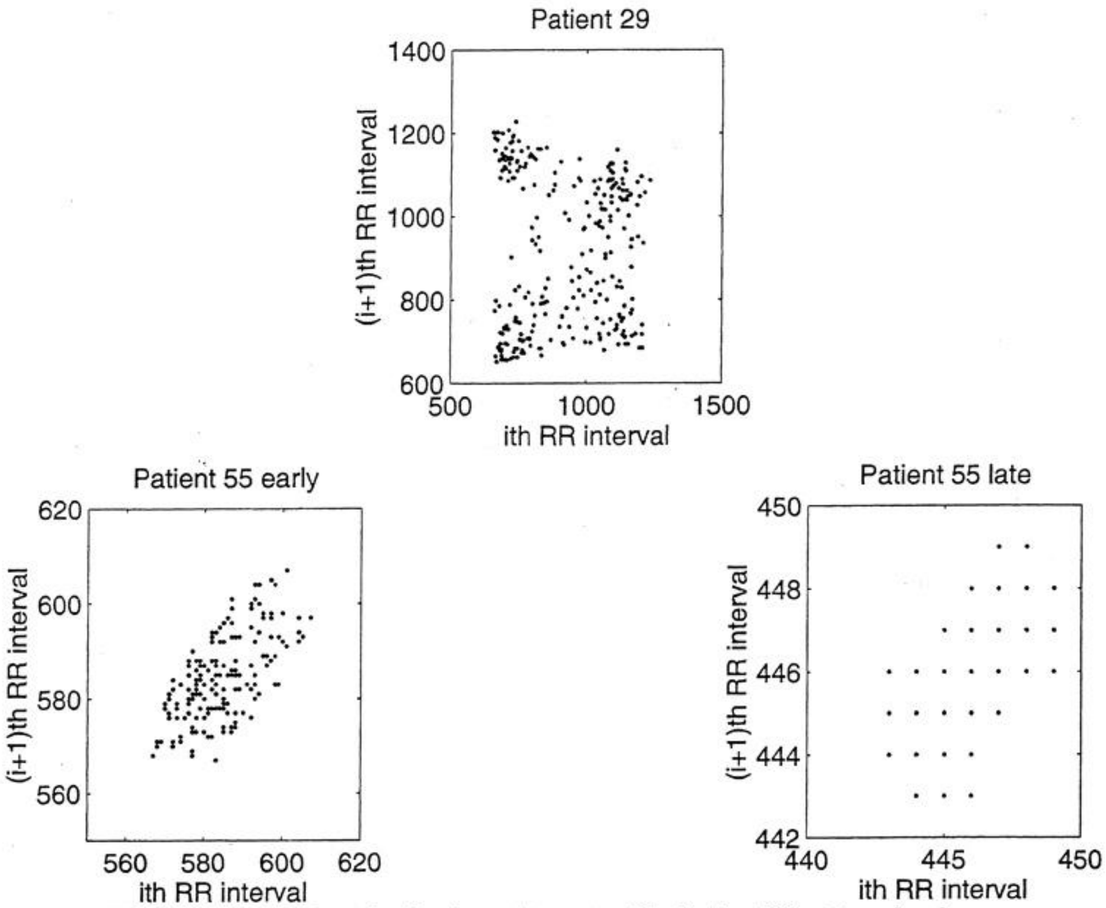


Figure 3 Lag 1 plots for the three data sets. The Patient 29 pattern has been described previously as a "complex pattern" (Woo et al. [1992]). Patient 55 early and late data plots would be classified as "torpedo patterns" by Woo et al. [1992].

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

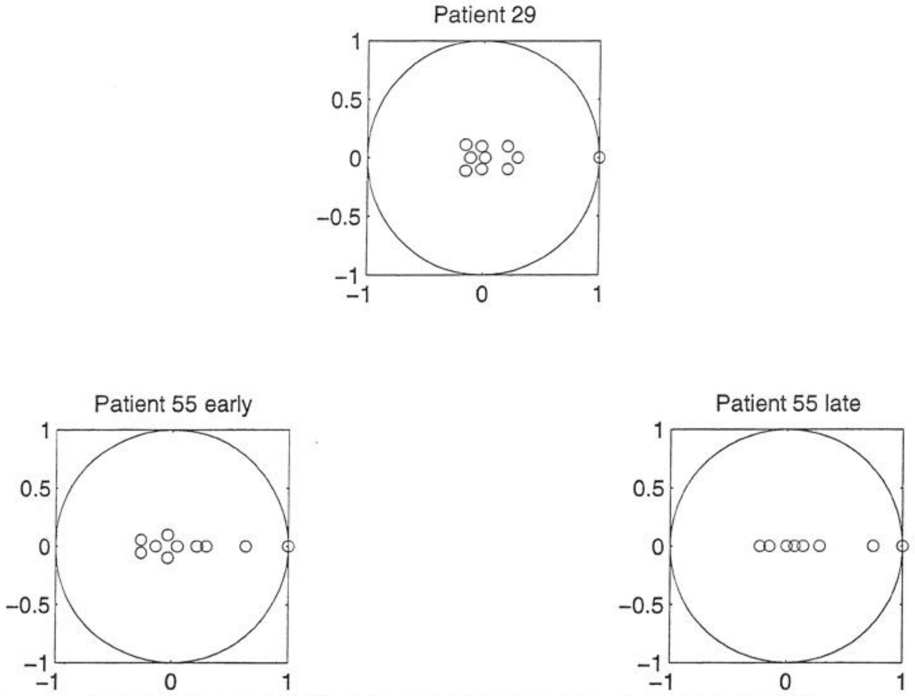


Figure 5 Eigenvalues for each of the three transition matrices are shown in relation to the unit circle. Besides the nearness of the "non-1" eigenvalues to the unit circle, notice the differences in the number of complex eigenvalues.

Basic idea in the HRV example

- Create an empirical Markov chain. Data is in the form of sequence of numbers and lag 1 maps indicate first order structure.
- Need to define the bin size corresponding to the length of the data set (usual number of bins used was 10). Then estimating transition probabilities to get a transition matrix. Note that many possible transitions are not observed.
- Transient aspects of the chain are of interest (not asymptotic behavior). Characterization of the dynamics (or the resulting matrix) is desired.
- Basic idea is to use properties of the matrix (such as eigenvalues) to distinguish between cases.

Advantages/disadvantages of the Markov chain model

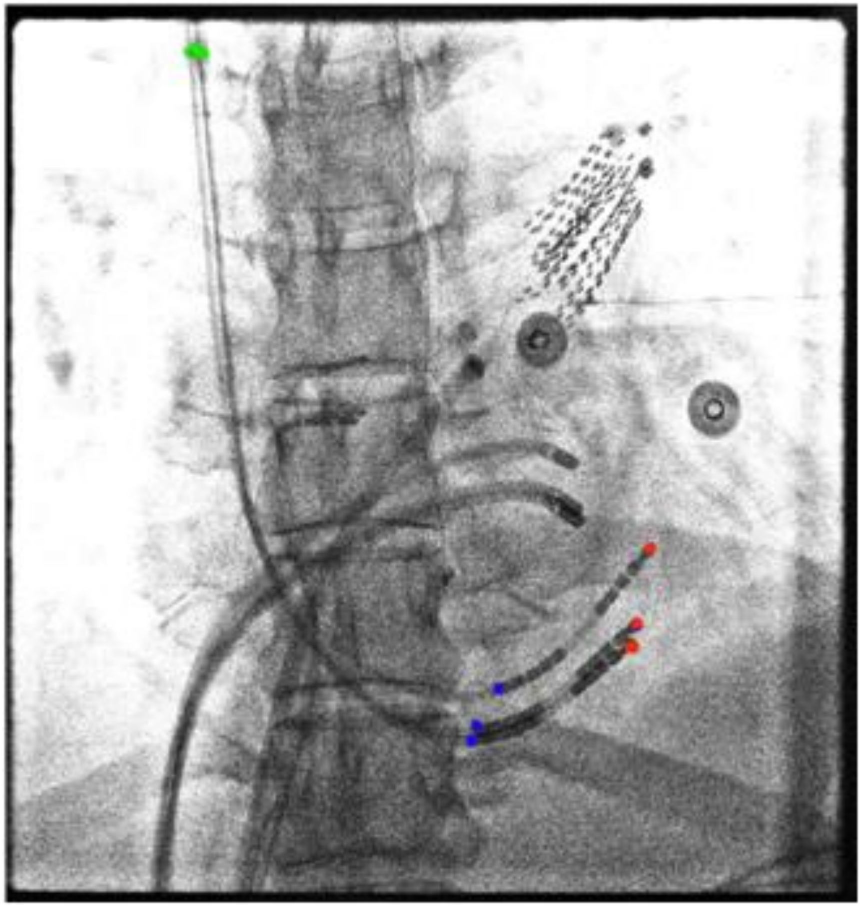
- Qualitative properties of the chain (e.g. eigenvalues or limit distribution) can be used to characterize the data set or identify changes in the data set.
- Simulation of the chain can result in an unlimited number of sample paths with the same dynamic behavior of the original set.
- Transition matrix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

Cardiac Image Registration

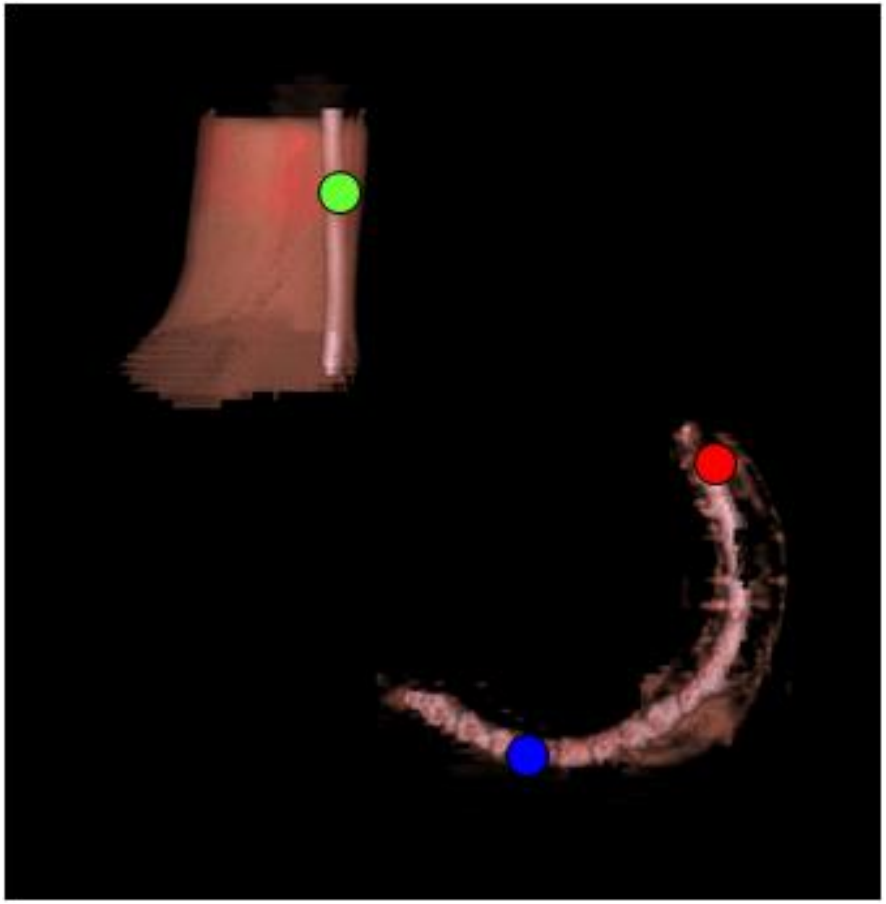
Shivani Ratnakumar

- 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



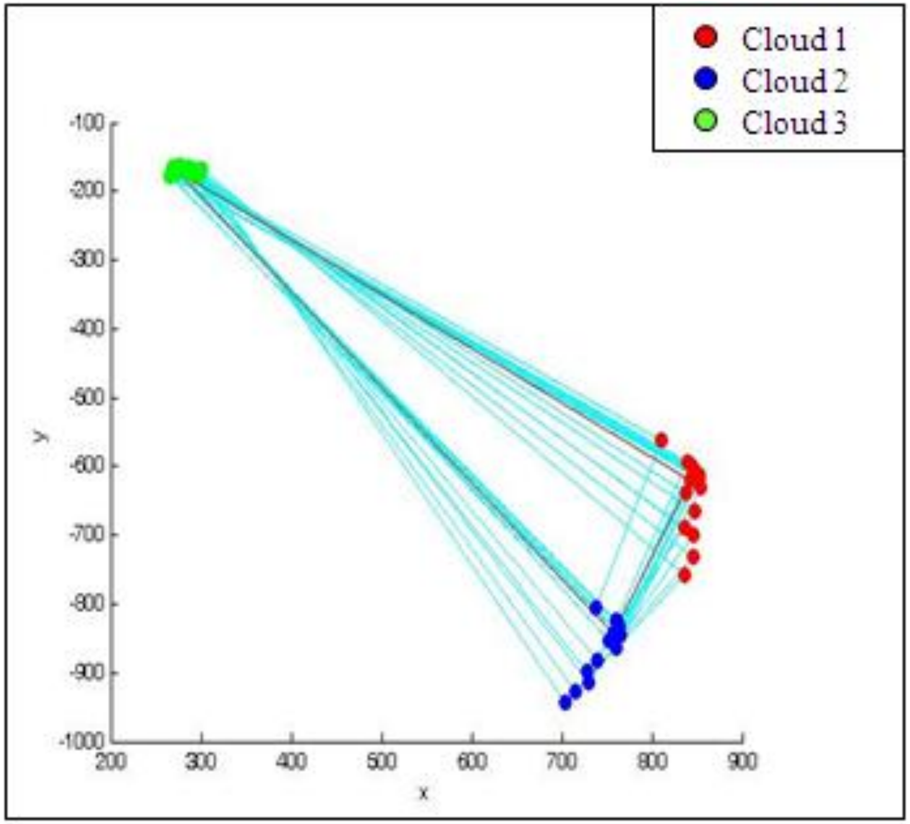
Three layered ECG gated fluoro frames



Segmented CT image that will be registered

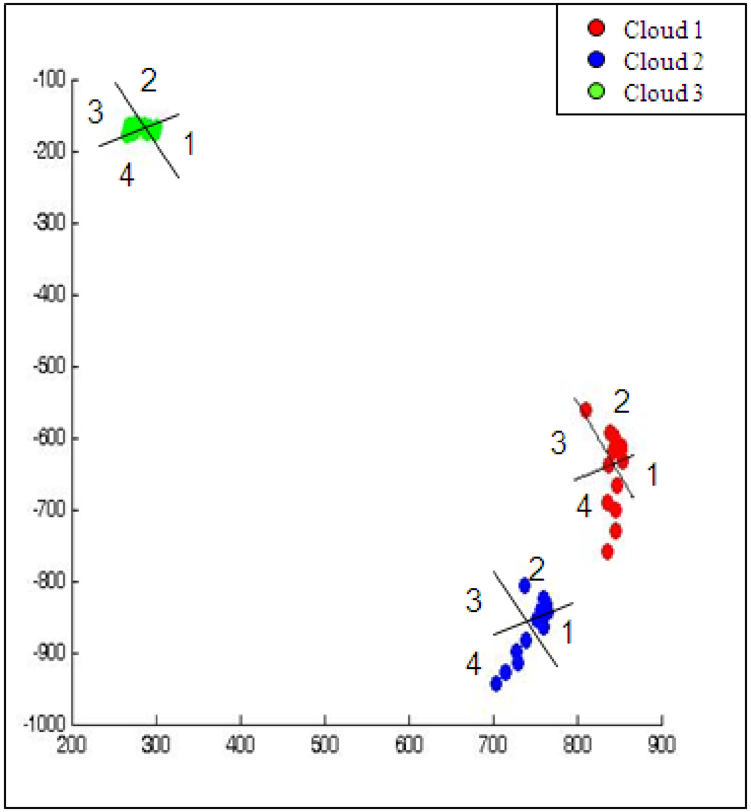
Concentrating on the dynamics

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



$$\begin{array}{ccc} 1 & 2 & 3 \\ \left[\begin{array}{ccc} (852,-618) & (766,-844) & (266,-177) \\ (810,-562) & (738,-806) & (268,-164) \\ \vdots & \vdots & \vdots \\ (848,-608) & (760,-836) & (273,-174) \end{array}\right] \end{array}$$

Discretize the state space



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

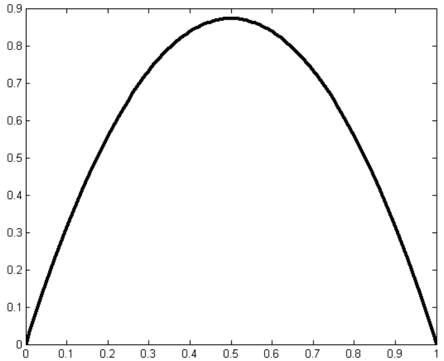
Sequence of coordinate points triplets can be written as a sequence of state triplets

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:



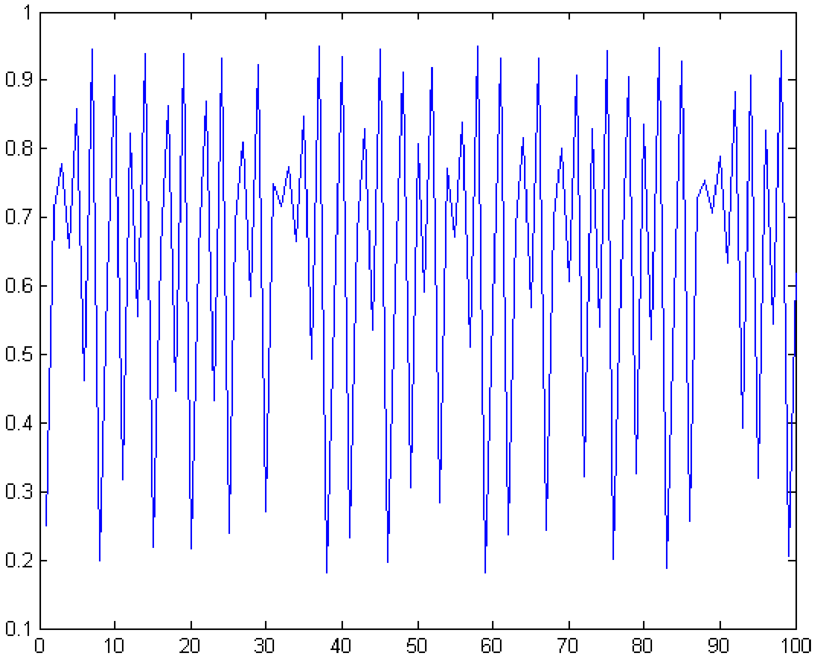
$$x_{n+1} = \mu x_n (1 - x_n) \quad 0 \le \mu \le 4$$

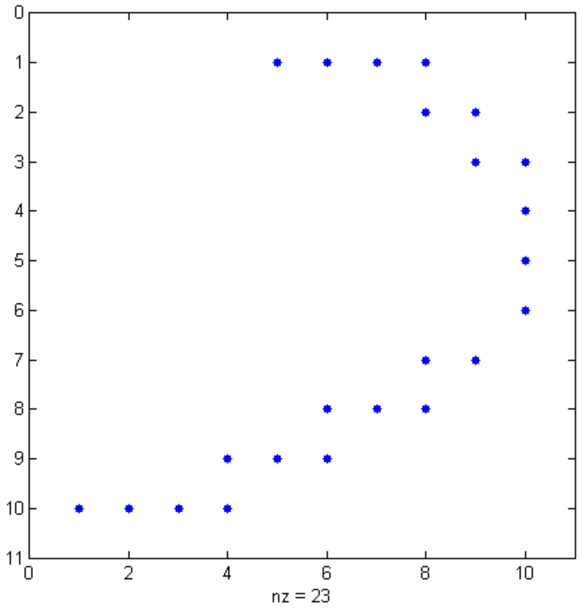
- Summary of behavior: $x = 0$ is always an equilibrium, stable if $\mu < 1$.

$x = 1 - \frac{1}{\mu}$ 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 $\mu = 3.8$

Bifurcation in Markov Chains

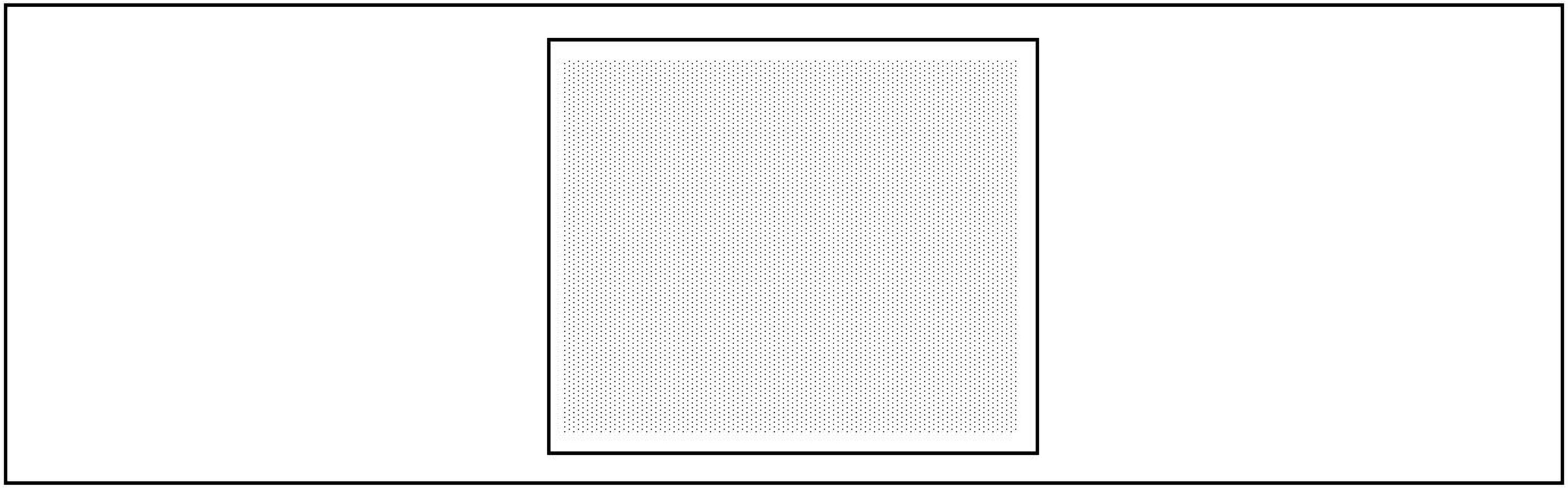
- One can use the empirical chain to detect bifurcations. In this case, the bifurcation at $\mu = 3$ to a period 2 point. This is done through the eigenvalues of the transition matrix.
- Facts: 1 is always an eigenvalue and others are on or inside the unit circle. If 1 is the only eigenvalue of magnitude 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**.

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.

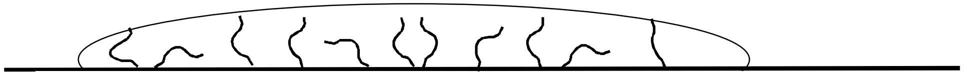
Spatial variation in the cDNA microarray

- cDNA microarray used to identify genes that are differentially under or over expressed in a sample (as seen through mRNA).



Some details of the process

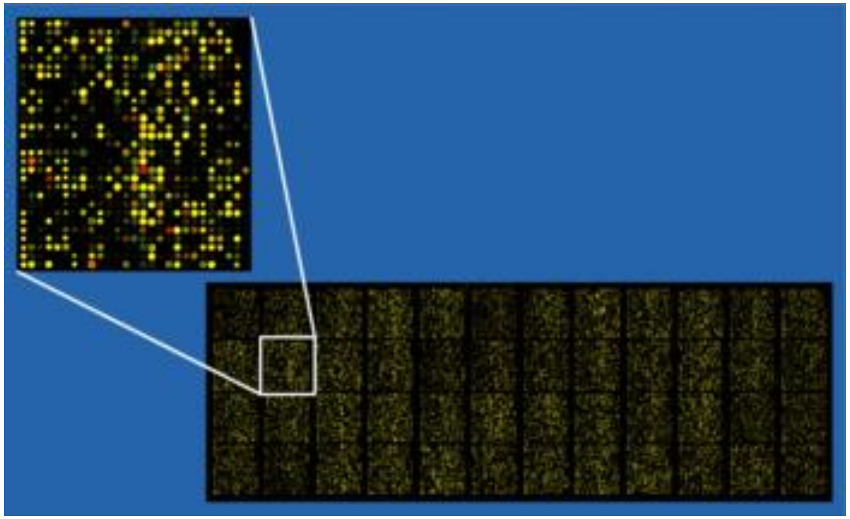
- the slide or chip is printed with a library of genes including those of special interest
- collect mRNA under two different conditions. Using RT and two different fluorescent dies, samples of labeled (“red” and “green”) DNA are produced.
- incubate the samples with the slide under a cover slip.



- scan the result to measure the amount of red and green fluorescence at each spot to measure the relative amount of mRNA present in the two samples.

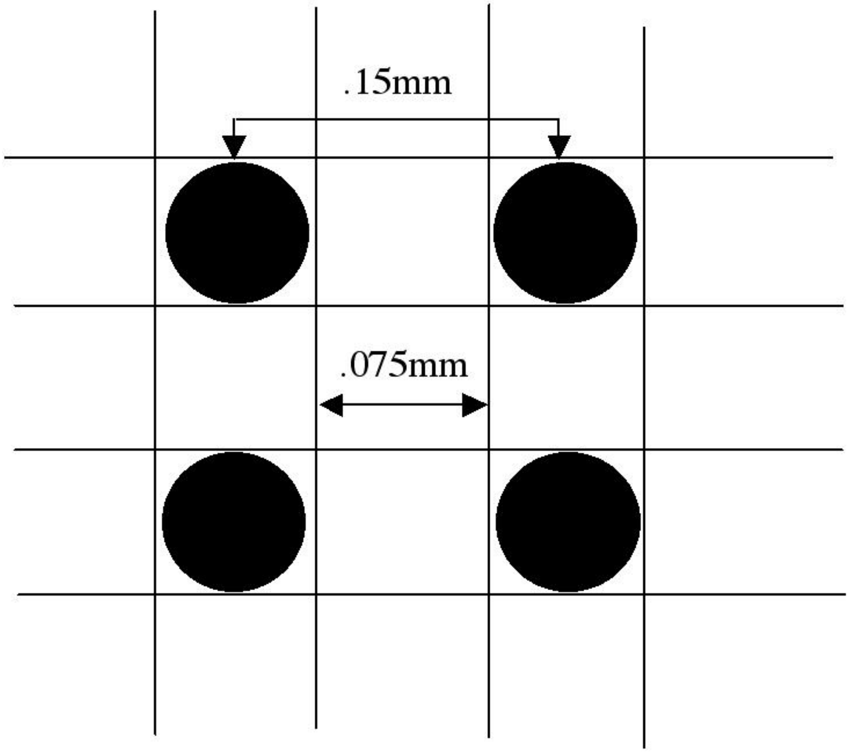
Motivation

- High variability -- both between replicates and within the same slide (with duplicated specificities in dots).
- Spatial variation in the brightness observed ("bright edges")
- Need to understand the proper normalization for this process

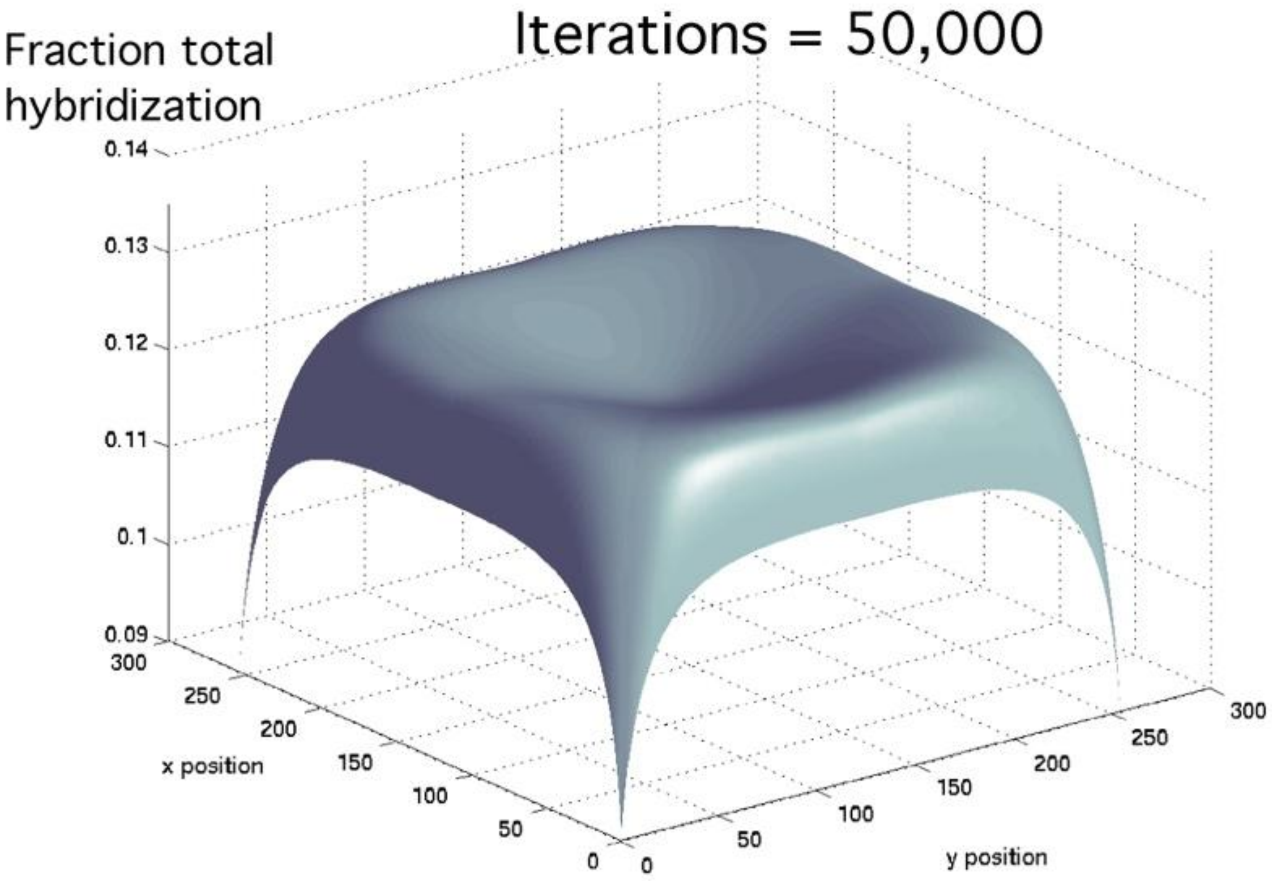


Construction of the MC chains

- Using the natural grid of positions on a slide, a Markov chain corresponding to **each of the 16,000 dots** is constructed. The goal being to compute the probability of absorption as a function of the transition number.
- The transition probabilities are based on the "taxi-cab" metric on the grid.



Results


Iterations = 50,000
Fraction total hybridization
0.14
0.13
0.12
0.11
0.1
0.09
x position
300
250
200
150
100
50
0
y position
0
50
100
150
200
250
300

Summary

- Empirical 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 especially the limit distribution and eigenvalues contain useful information on the process.

References

- Merrill, S.J. & Cochran, J.R. (1997) Markov chain methods in the analysis of heart rate variability, *Fields Inst Comm* **11**:241-252.
- S.J. Merrill, S. Nelson, and C.A. Struble, Spatial dependence of hybridization in the cDNA microarray, *Can. Appl. Math. Quarterly* **11** (2003), 321-337.
- 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.*
- Vincent, K., Merrill, S., Struble, C. and Hecox, K. (2009) Markov chain analysis of pediatric ictal EEGs, Abstract for the American Epilepsy Society Annual Meeting (AES2009) Boston Dec. 2009.