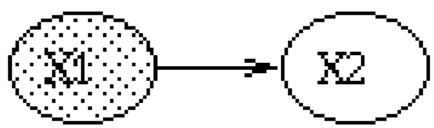
Modeling Sequential Processes

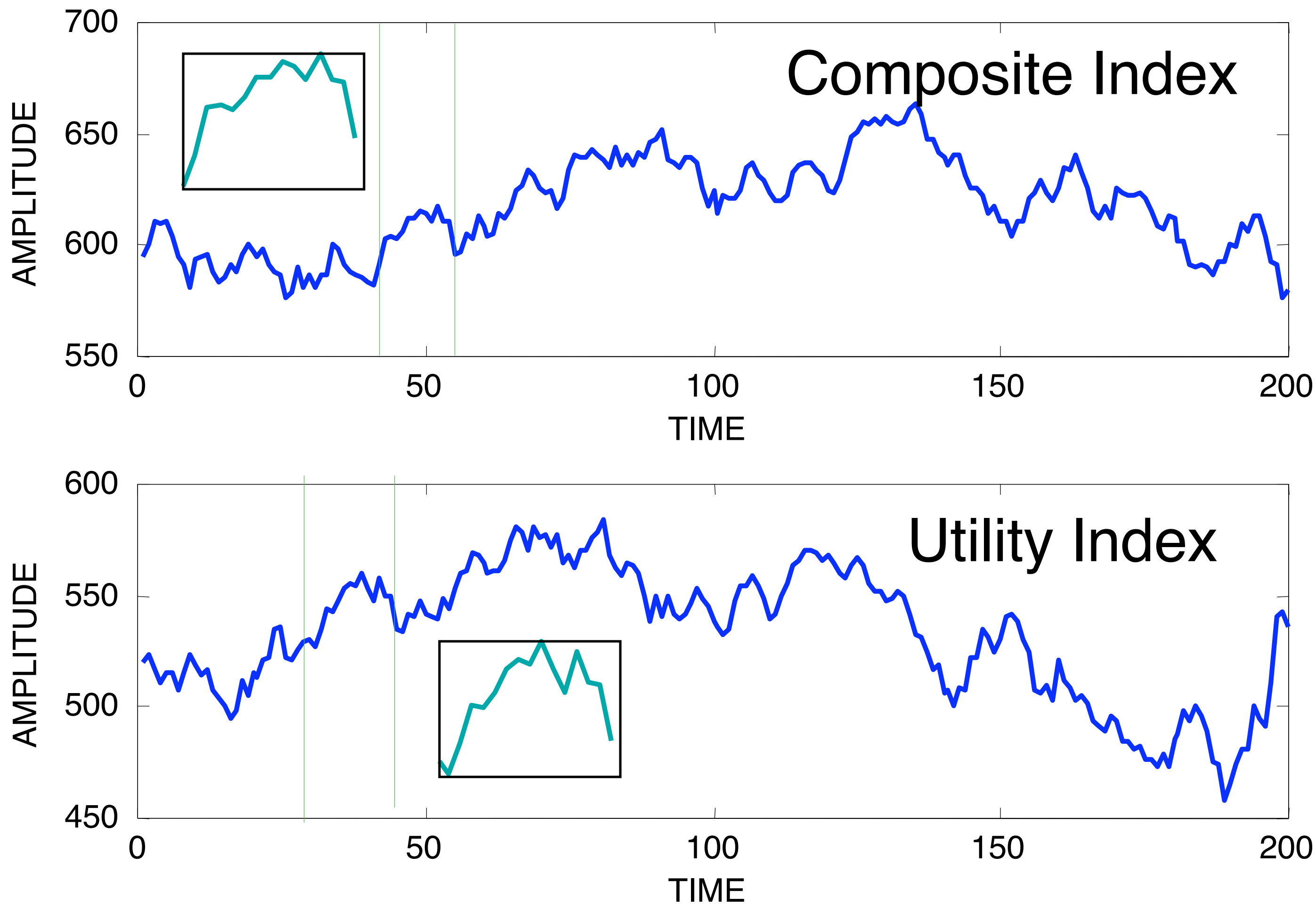
Simple Sequential Processes

- Sequences of Events: State Dynamics
- Sequences of Responses
- Sequences of Decisions

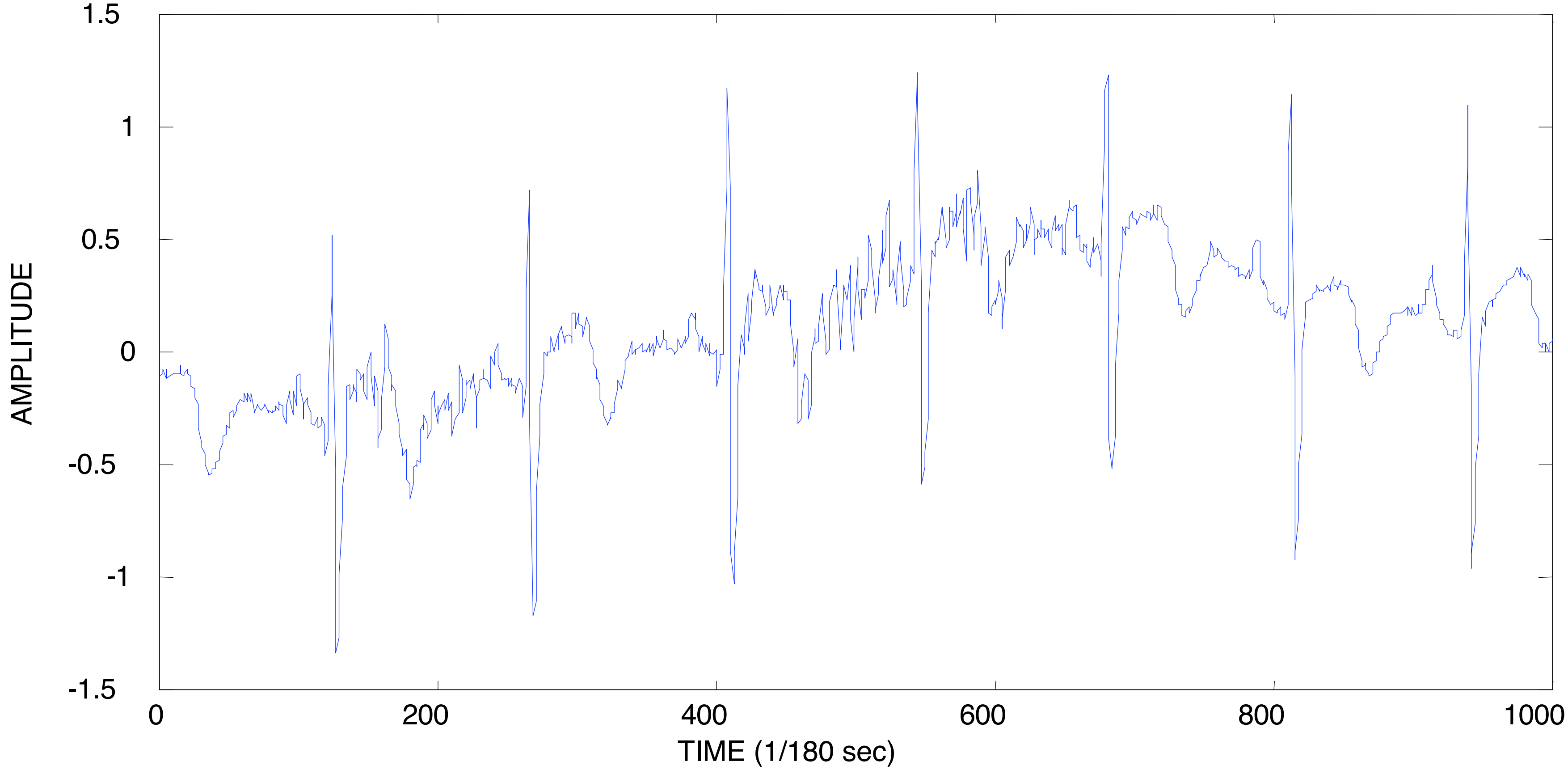


Auto Regressive model AR(1)

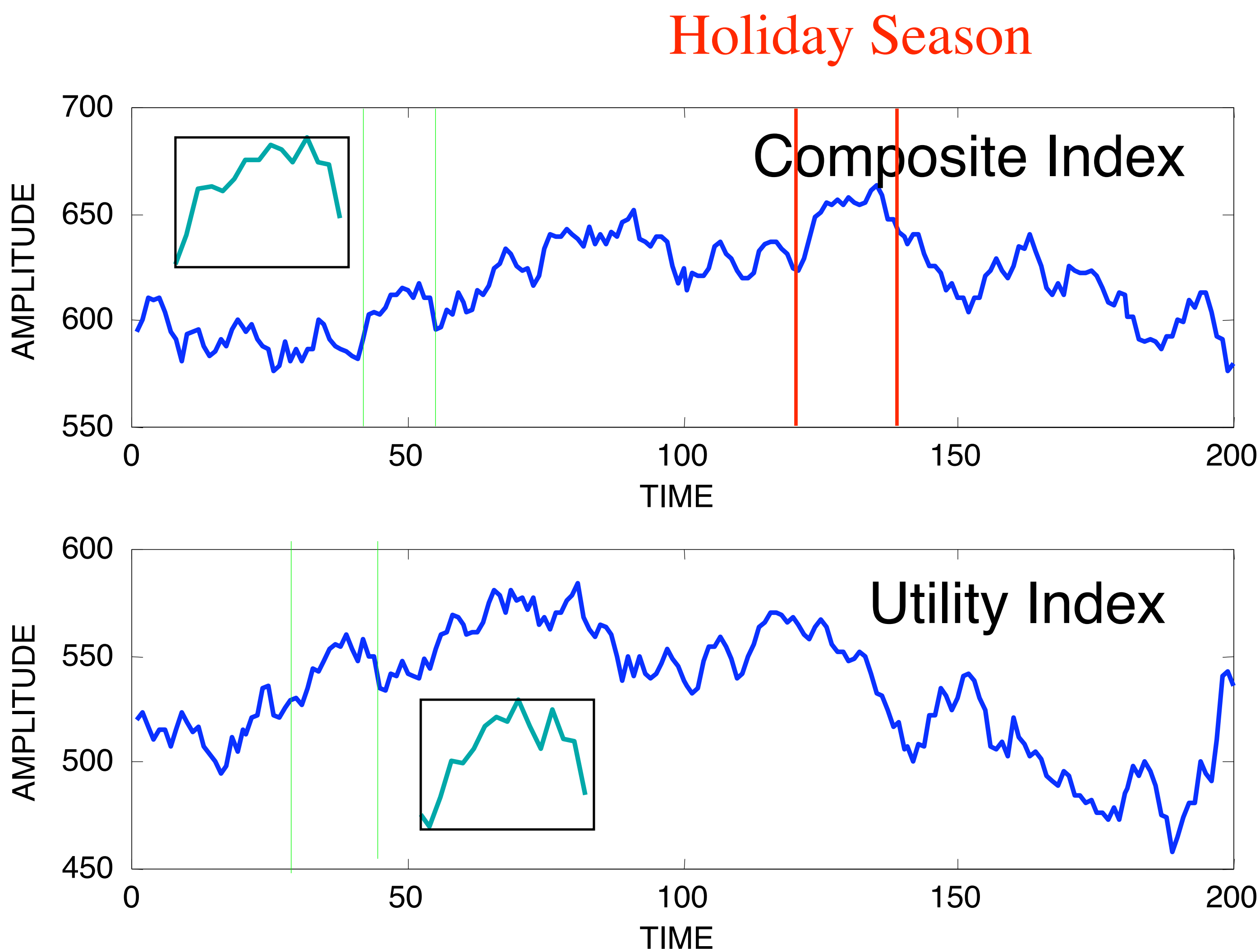
Time Series Data



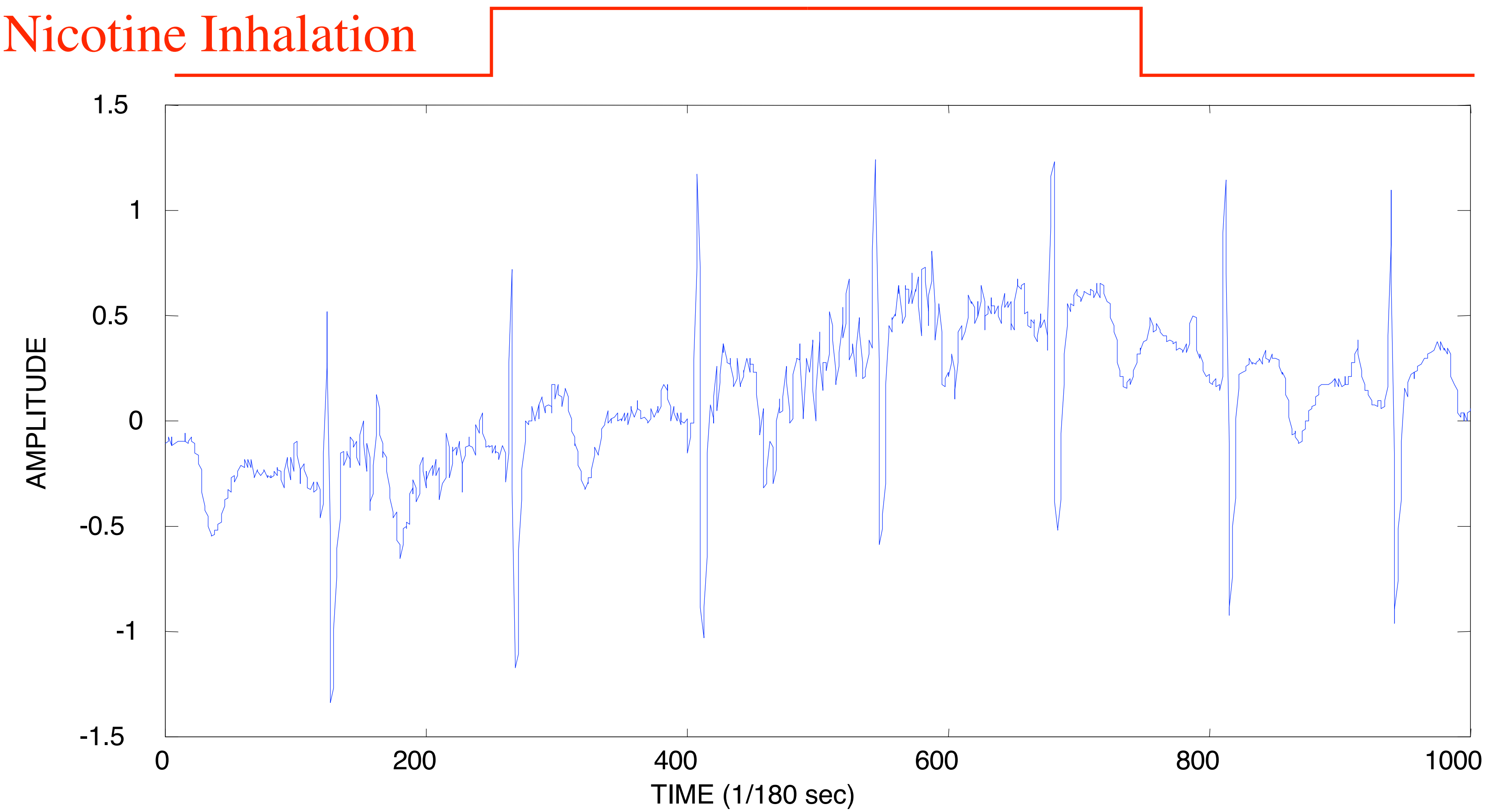
ECG Data



Time Series Data with Hidden state



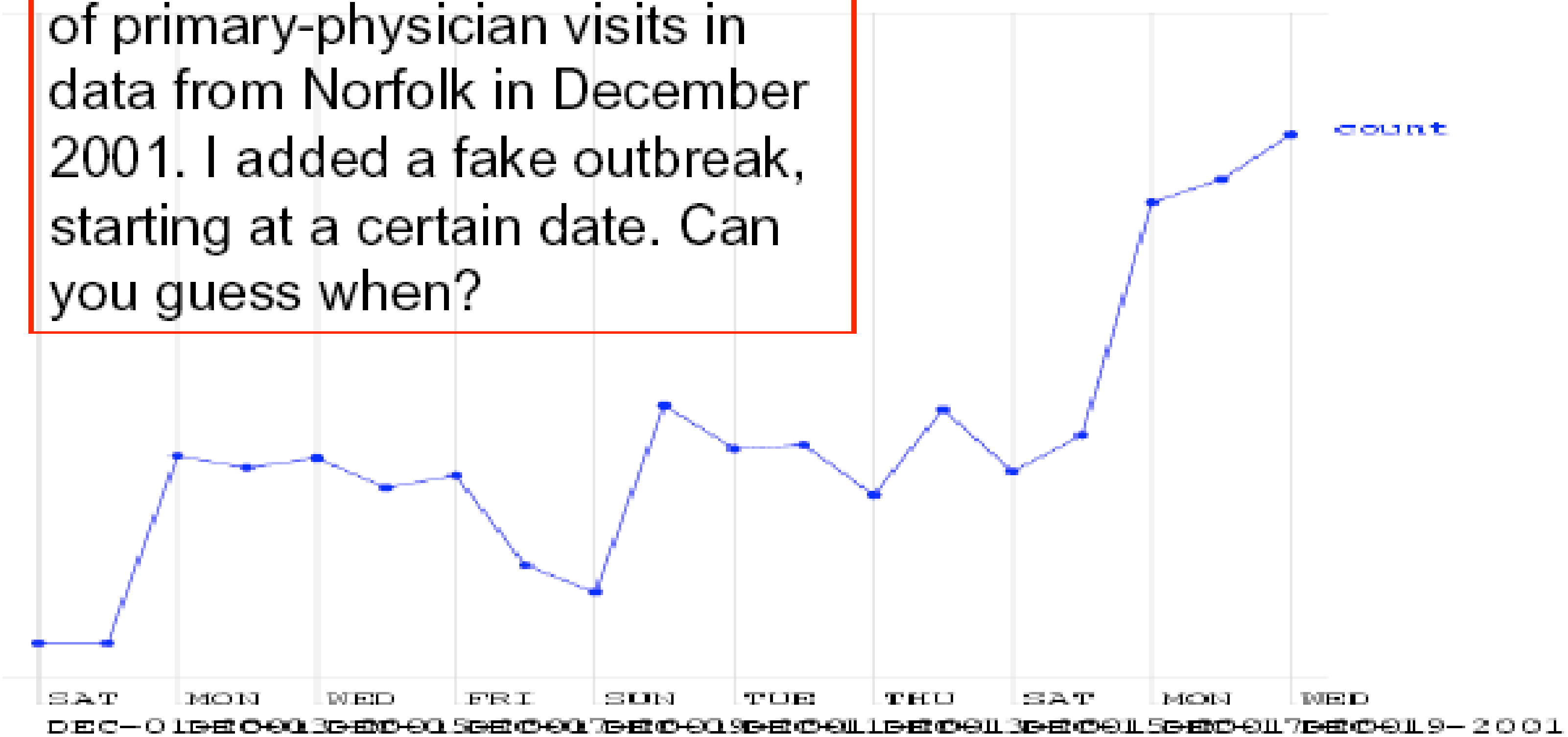
ECG Data with Hidden State



Anomaly detection

(When) is there an anomaly?

This is a time series of counts of primary-physician visits in data from Norfolk in December 2001. I added a fake outbreak, starting at a certain date. Can you guess when?

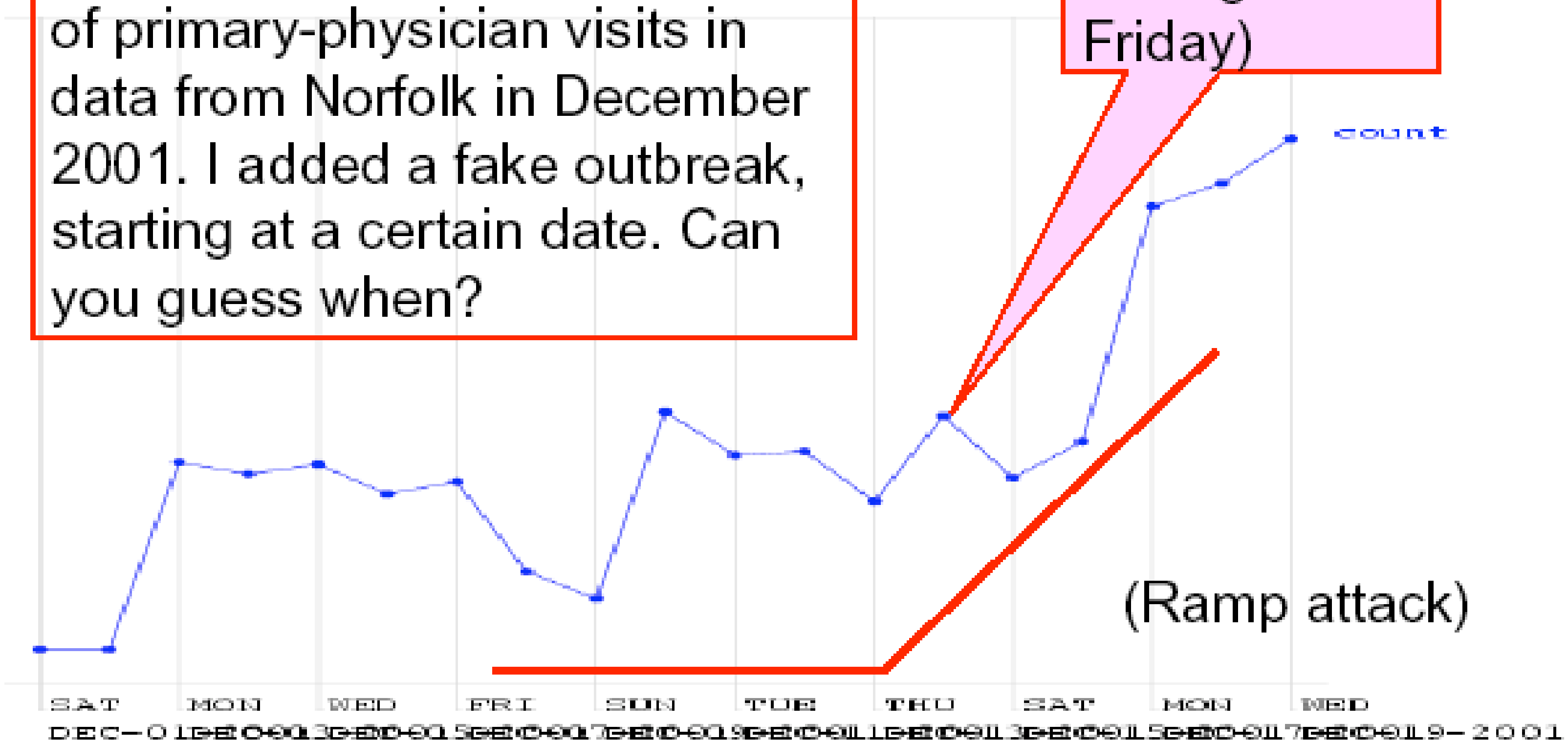


Did you get it right?

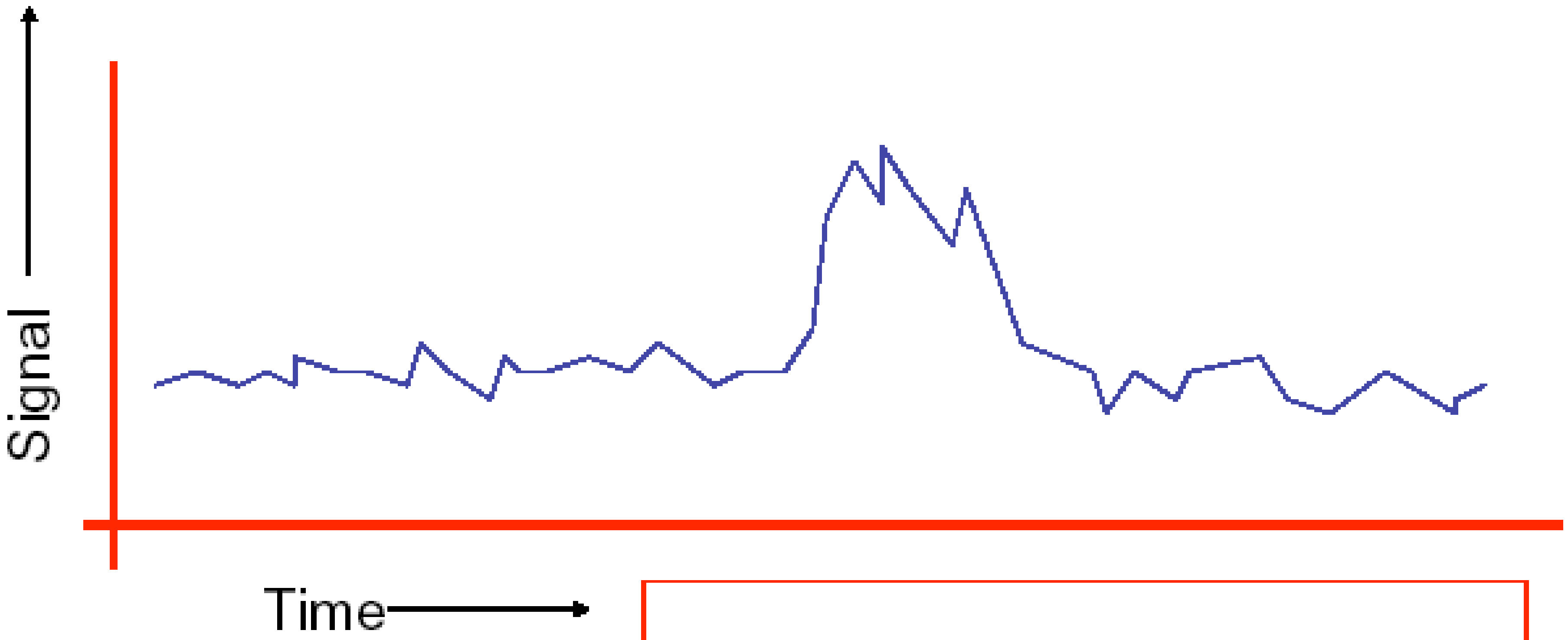
(When) is there an anomaly?

This is a time series of counts of primary-physician visits in data from Norfolk in December 2001. I added a fake outbreak, starting at a certain date. Can you guess when?

Here (much too high for a Friday)



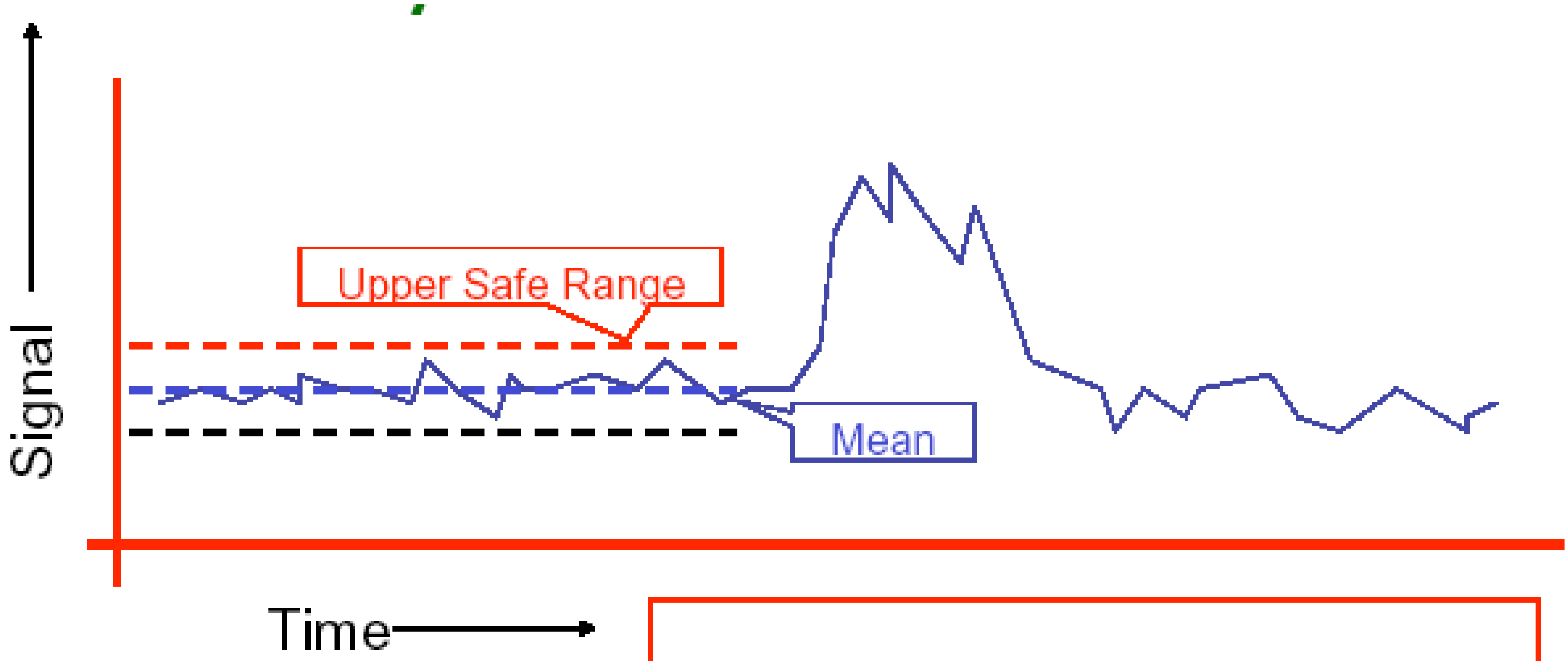
An easy case



Dealt with by Statistical Quality Control

Record the mean and standard deviation up the the current time.

Signal an alarm if we go outside 3 sigmas

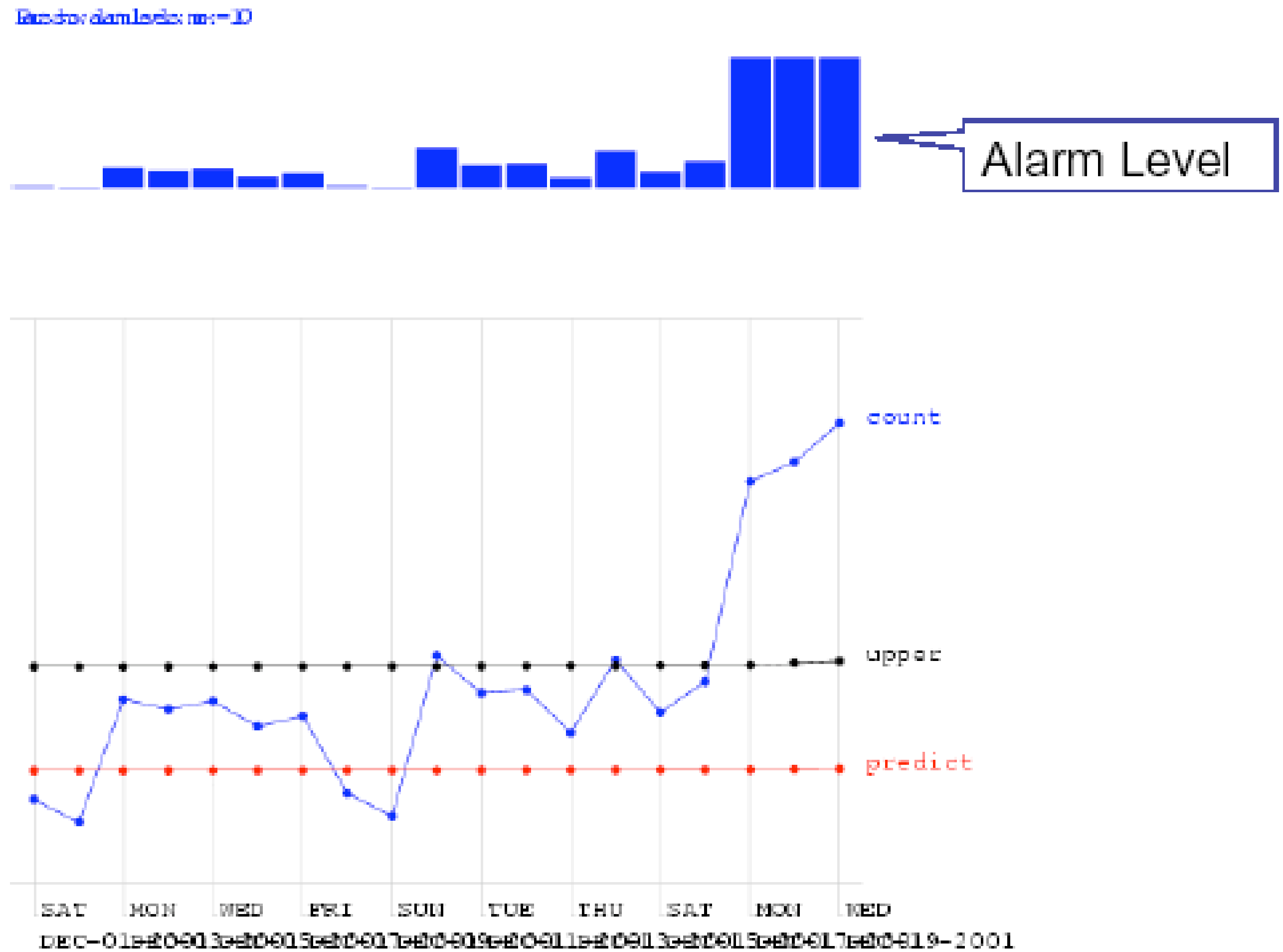


Dealt with by Statistical Quality Control

Record the mean and standard deviation up the the current time.

Signal an alarm if we go outside 3 sigmas

Control Charts on the Norfolk Data



Time series Inference Tasks

- Structure detection

  100010001000100010001000

  - Find a simple model for behavior
  - Choose between models for behavior
- Prediction

  10101001010?
- Anomaly detection

  100010001010

  a b c d e f q h i j k l m u o p

- *How do we solve these problems?*

Use Time series models:

$s_T = f(\text{past}) + \text{noise}$

where

s_T = state at time T

Future predictable from the past:

$s_T = f(\{s_1, s_2, s_3, \ldots, s_{T-1}\}, t)$

$s_T = f(\text{past}) + \text{noise}$

Time Series Data

- You are given a collection of labelled points in some order: { (y_1,x_1), (y_2, x_2),…., (y_N, x_N) }. (e.g. x_i are category labels, y_i are measurements).
- In time series data, *independence is violated.*

$$p(y_1, y_2, \ldots, y_n, x_1, x_2, \ldots, x_n) \neq p(y_1 \mid x_1)\, p(y_2 \mid x_2)\, p(y_n \mid x_n)$$

- In time series data, *order matters.*

Basic decomposition for any distribution

$$p(y_1, x_1, x_2, y_2, \ldots, y_n, x_n) = p(y_n, x_n \mid x_{n-1}, y_{n-1}, \ldots, y_1, x_1) \cdot$$
$$p(x_{n-1}, y_{n-1} \mid x_{n-2}, y_{n-2}, \ldots, y_1, x_1) \cdots p(y_1 \mid x_1)$$

Markov Assumption

$$p(y_1, x_1, x_2, y_2, \ldots, y_n, x_n) = p(y_n, x_n \mid x_{n-1}, y_{n-1})\, p(x_{n-1}, y_{n-1} \mid x_{n-2}, y_{n-2}) \cdots p(y_1 \mid x_1)$$

Describing Sequential States

$$p(X_n = S_i | X_{n-1} \ldots X_{n-j} = s) = a.$$

S is discrete or continuous

Independence

$$p(X_n = S_i | X_{n-1} \ldots X_{n-j} = s) = p(X_n = S_i);$$

Stationarity

$$p(X_n = S_i | X_{n-1} \ldots X_{n-j} = s) = p(X_m = S_i | X_{m-1} \ldots X_{m-j} = s);$$

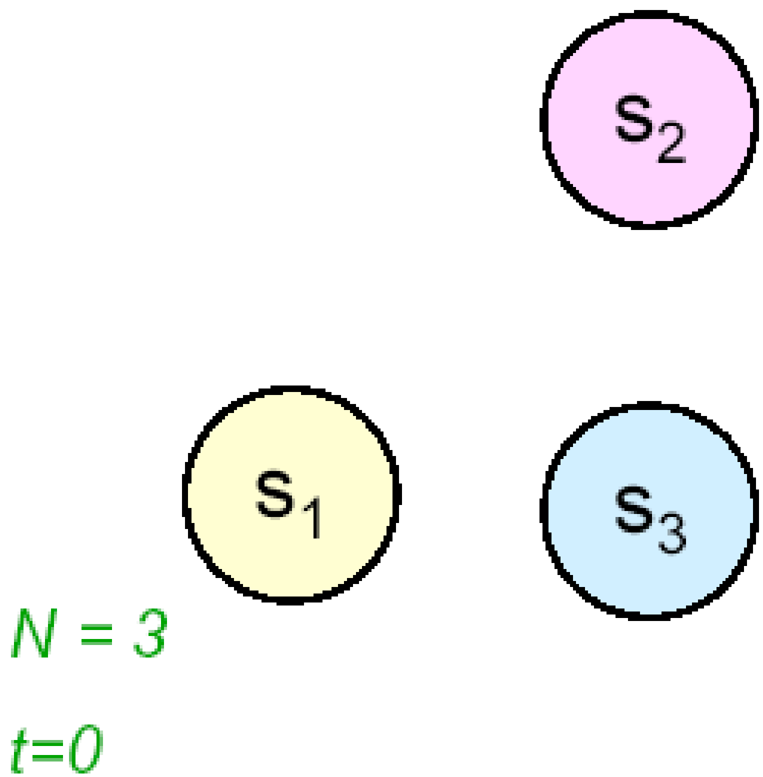
Markov

$$p(X_n = S_i | X_{n-1} \ldots X_{n-j} = s) = p(X_n = S_i | X_{n-1} = S_j);$$

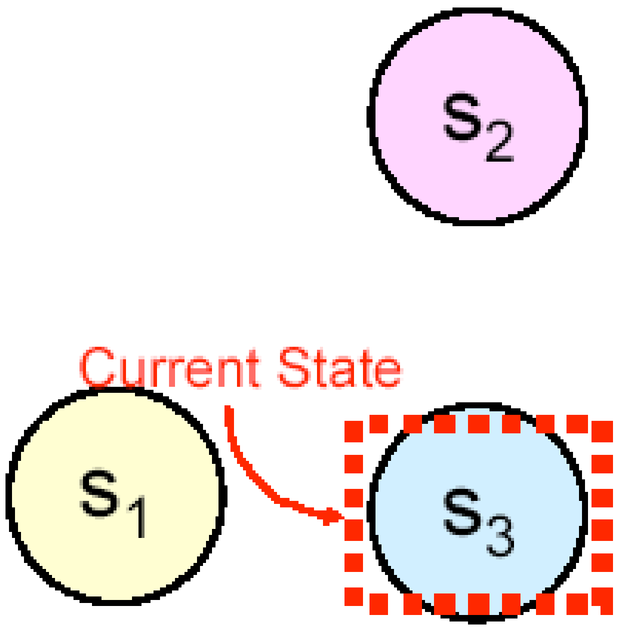
A Markov System

Has N states, called s_1, s_2 .. s_N

There are discrete timesteps, $t=0$, $t=1$, …



A Markov System



$N = 3$

$t=0$

$q_t=q_0=s_3$

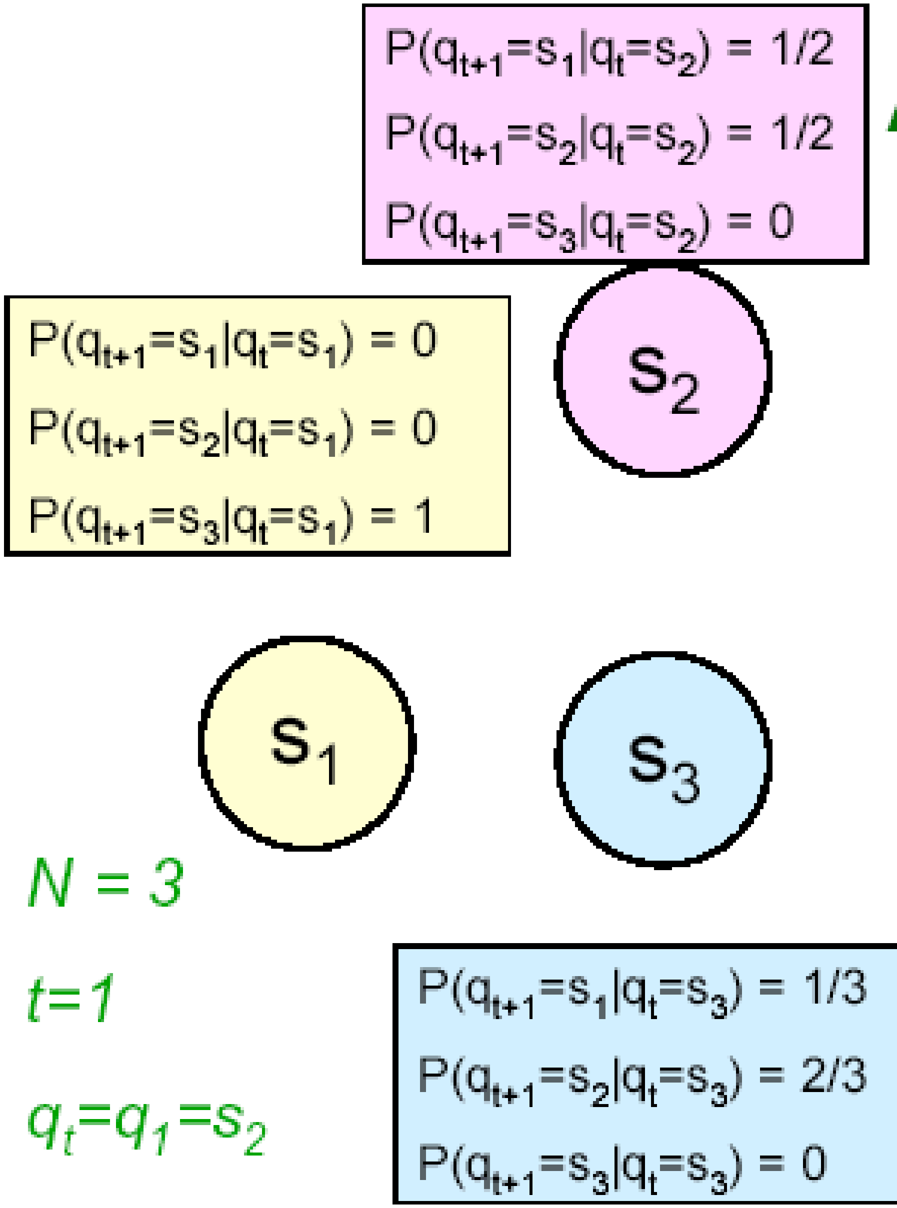
Has N states, called s_1, s_2 .. s_N

There are discrete timesteps, $t=0$, $t=1$, …

On the t'th timestep the system is in exactly one of the available states. Call it q_t

Note: $q_t \in \{s_1, s_2 .. s_N\}$

A Markov System



Has N states, called $s_1, s_2 .. s_N$

There are discrete timesteps, $t=0, t=1, \ldots$

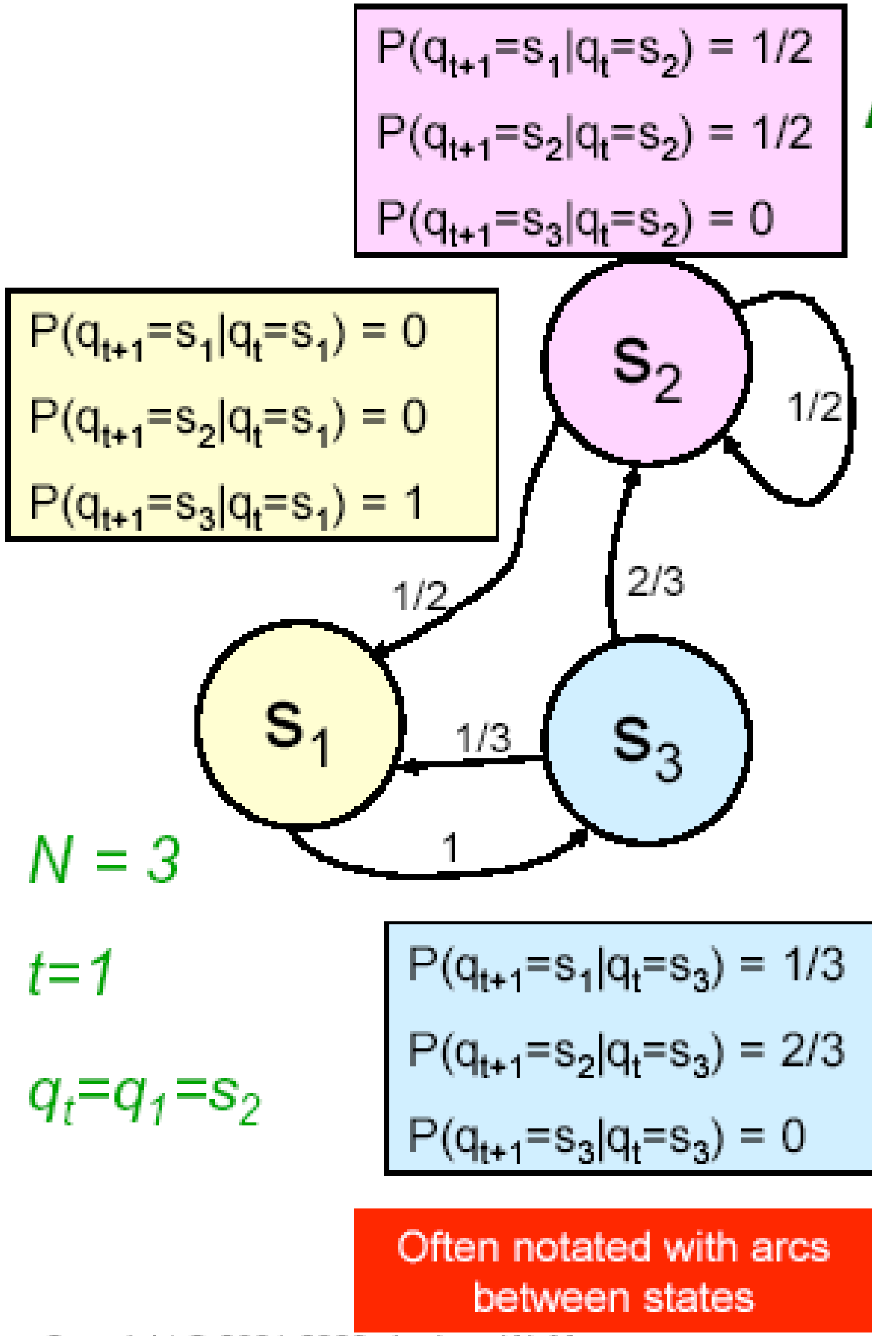
On the t'th timestep the system is in exactly one of the available states. Call it q_t

Note: $q_t \in \{s_1, s_2 .. s_N\}$

Between each timestep, the next state is chosen randomly.

The current state determines the probability distribution for the next state.

A Markov System



Has N states, called s_1, s_2 .. s_N

There are discrete timesteps, $t=0$, $t=1$, …

On the t'th timestep the system is in exactly one of the available states. Call it q_t

Note: $q_t \in \{s_1, s_2 .. s_N\}$

Between each timestep, the next state is chosen randomly.

The current state determines the probability distribution for the next state.

Markov Property

$P(q_{t+1}=s_1|q_t=s_2) = 1/2$

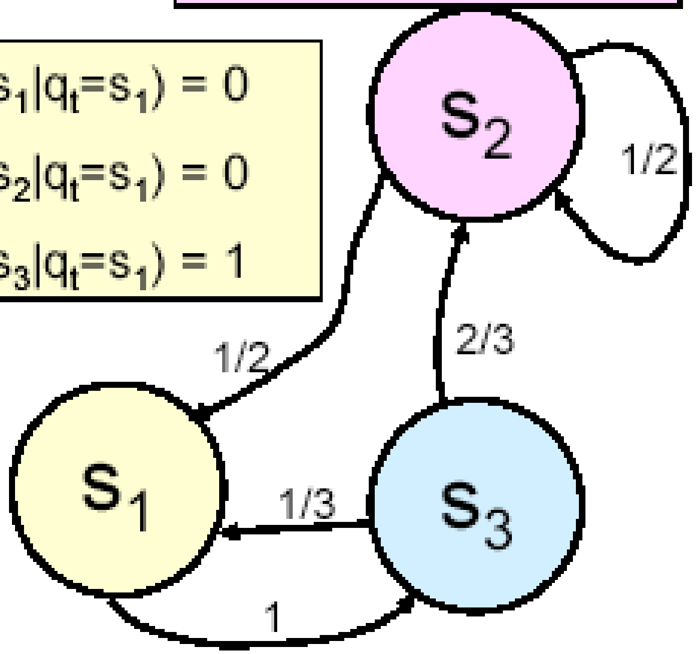
$P(q_{t+1}=s_2|q_t=s_2) = 1/2$

$P(q_{t+1}=s_3|q_t=s_2) = 0$

$P(q_{t+1}=s_1|q_t=s_1) = 0$

$P(q_{t+1}=s_2|q_t=s_1) = 0$

$P(q_{t+1}=s_3|q_t=s_1) = 1$



$N = 3$

$t=1$

$q_t=q_1=s_2$

$P(q_{t+1}=s_1|q_t=s_3) = 1/3$

$P(q_{t+1}=s_2|q_t=s_3) = 2/3$

$P(q_{t+1}=s_3|q_t=s_3) = 0$

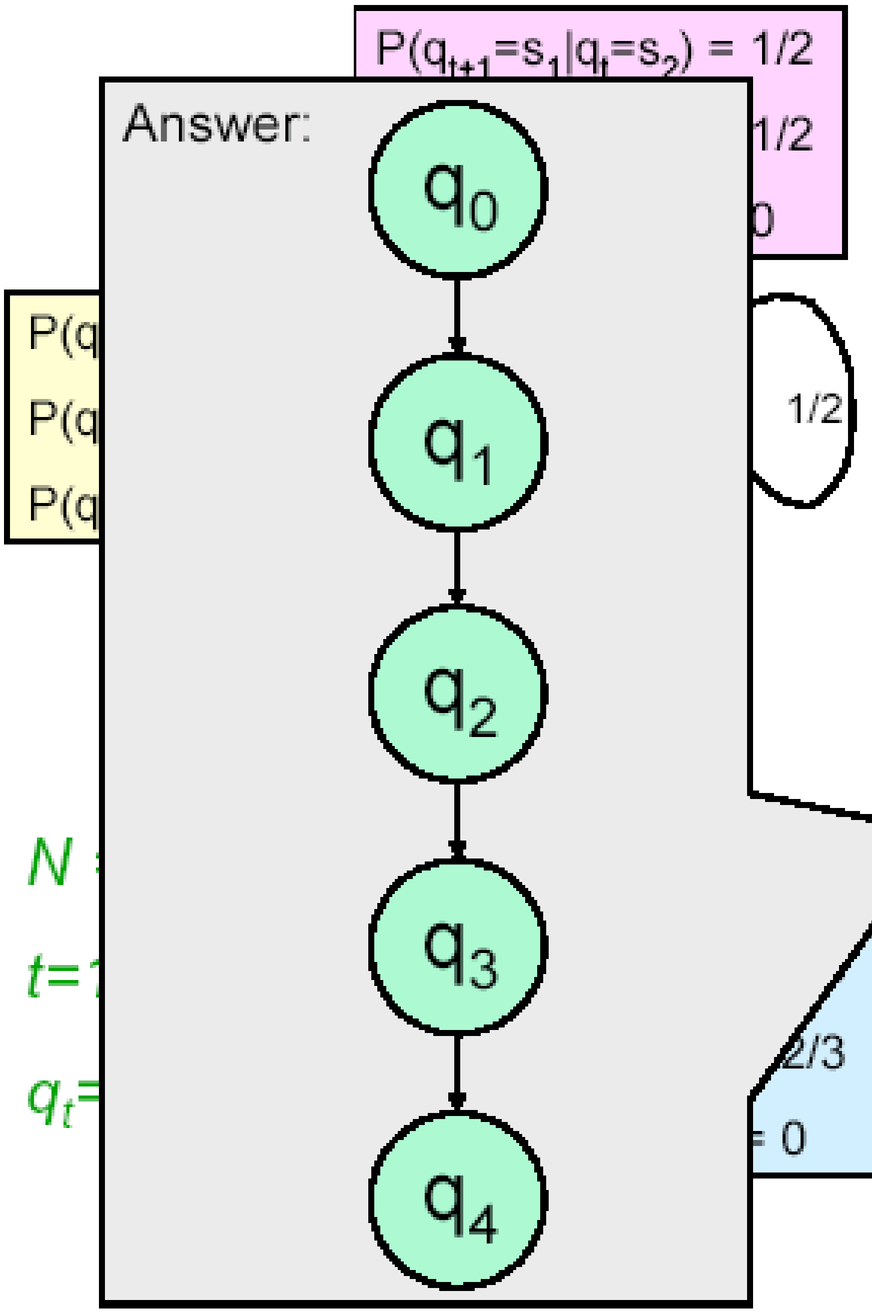
q_{t+1} is conditionally independent of $\{ q_{t-1}, q_{t-2}, \ldots q_1, q_0 \}$ given q_t.

In other words:

$P(q_{t+1} = s_j | q_t = s_i) =$

$P(q_{t+1} = s_j | q_t = s_i ,\text{any earlier history})$

Question: what would be the best Bayes Net structure to represent the Joint Distribution of ($q_0, q_1, \ldots q_3, q_4$)?



Markov Property

q_{t+1} is conditionally independent of $\{ q_{t-1}, q_{t-2}, \ldots q_1, q_0 \}$ given q_t.

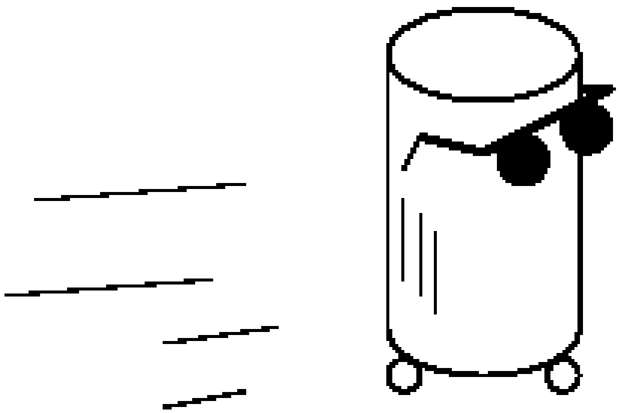
In other words:

$P(q_{t+1} = s_j \mid q_t = s_i) =$

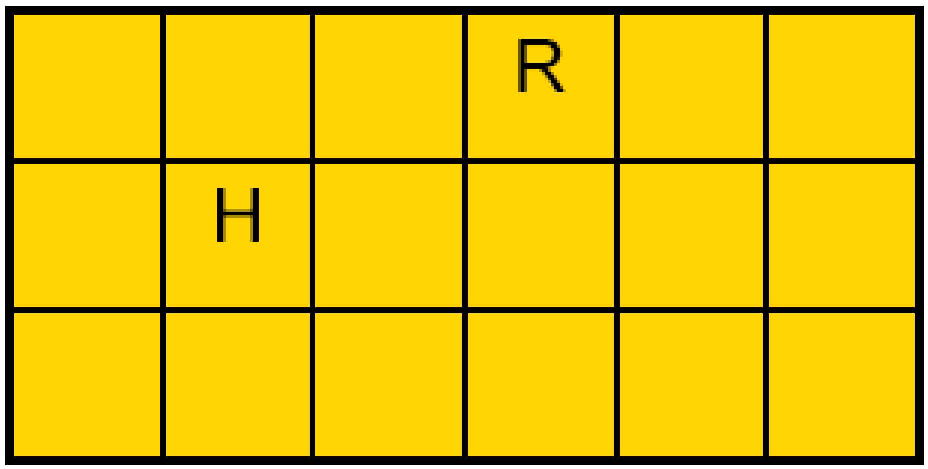
$P(q_{t+1} = s_j \mid q_t = s_i, \text{any earlier history})$

Question: what would be the best Bayes Net structure to represent the Joint Distribution of (q_0, q_1, q_2, q_3, q_4)?

A Blind Robot



A human and a robot wander around randomly on a grid…

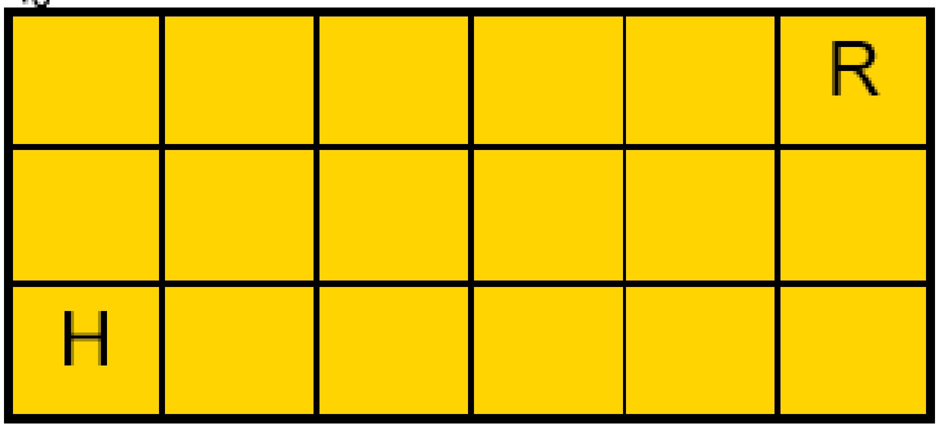


STATE q = ⟨ Location of Robot, Location of Human ⟩

Note: N (num. states) = 18 * 18 = 324

Dynamics of System

q_0 =

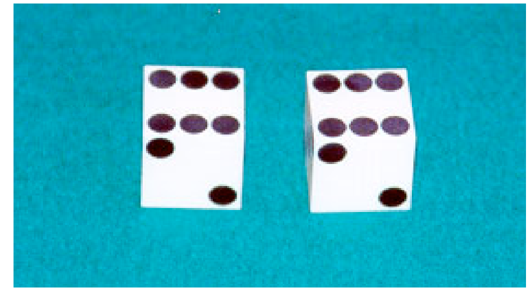


Each timestep the human moves randomly to an adjacent cell. And Robot also moves randomly to an adjacent cell.

Typical Questions:

- "What's the expected time until the human is crushed like a bug?"
- "What's the probability that the robot will hit the left wall before it hits the human?"
- "What's the probability Robot crushes human on next time step?"

Example: The Dishonest Casino



A casino has two dice:

- Fair die
  P(1) = P(2) = P(3) = P(5) = P(6) = 1/6
- Loaded die
  P(1) = P(2) = P(3) = P(5) = 1/10
  P(6) = 1/2

Casino player switches back-&-forth between fair and loaded die once every 20 turns

Game:

1. You bet $1
2. You roll (always with a fair die)
3. Casino player rolls (maybe with fair die, maybe with loaded die)
4. Highest number wins $2

Problem 1: Evaluation

GIVEN

A sequence of rolls by the casino player

1245526462146146136136661664661636616366163616515615115146123562344

QUESTION

How likely is this sequence, given our model of how the casino works?

This is the **EVALUATION** problem

Problem 2 – Decoding

GIVEN

A sequence of rolls by the casino player

1245526462146146136136661664661636616366163616515615115146123562344

QUESTION

What portion of the sequence was generated with the fair die, and what portion with the loaded die?

This is the **DECODING** question

Problem 3 – Learning

GIVEN

A sequence of rolls by the casino player

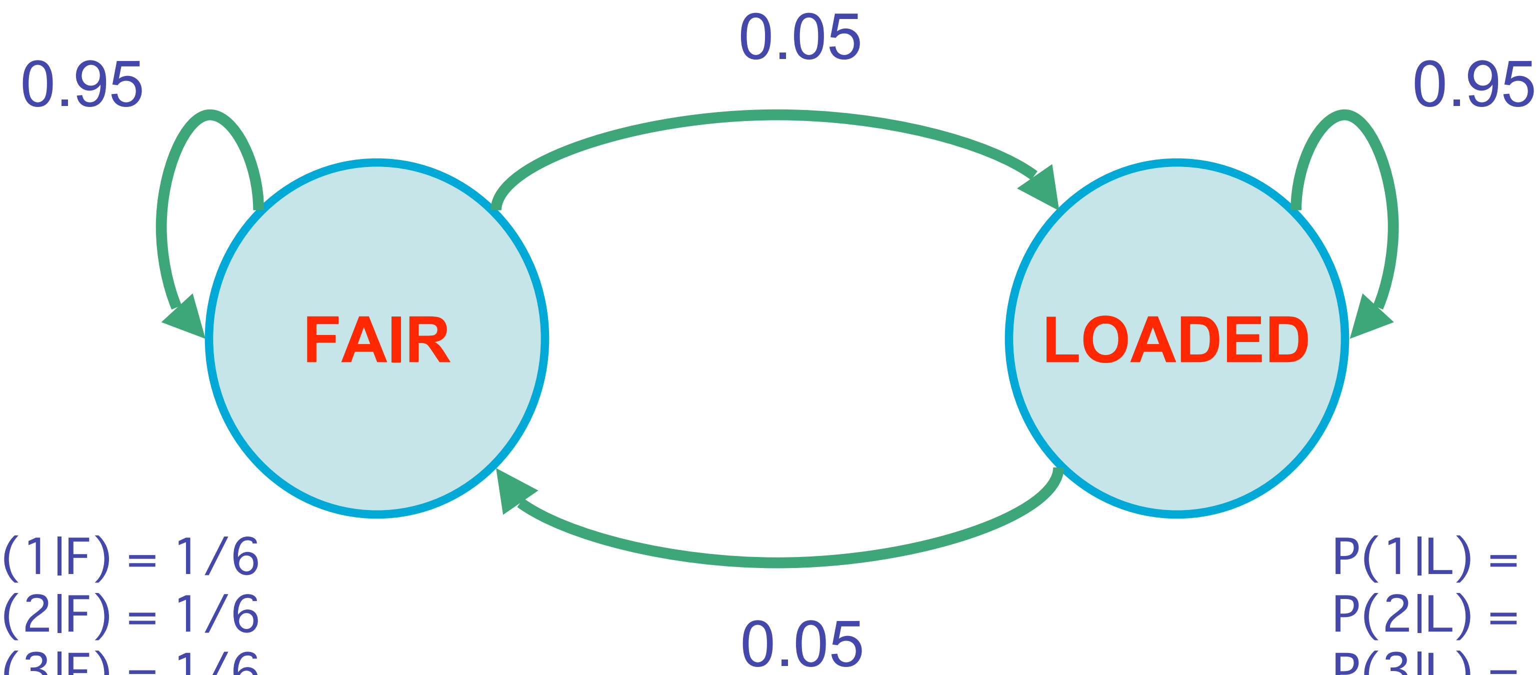
12455264621461461361366616646616366163661636165156151151461235623644

QUESTION

How "loaded" is the loaded die? How "fair" is the fair die? How often does the casino player change from fair to loaded, and back?

This is the **LEARNING** question

The dishonest casino model



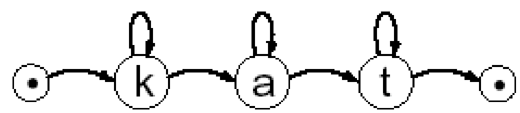
P(1|F) = 1/6
P(2|F) = 1/6
P(3|F) = 1/6
P(4|F) = 1/6
P(5|F) = 1/6
P(6|F) = 1/6

P(1|L) = 1/10
P(2|L) = 1/10
P(3|L) = 1/10
P(4|L) = 1/10
P(5|L) = 1/10
P(6|L) = 1/2

- **HMM is specified by:**
  - transition probabilities $p(q_n^j | q_{n-1}^i) \equiv a_{ij}$
  - (initial state probabilities $p(q_1^i) \equiv \pi_i$)
  - emission distributions $p(x | q^i) \equiv b_i(x)$

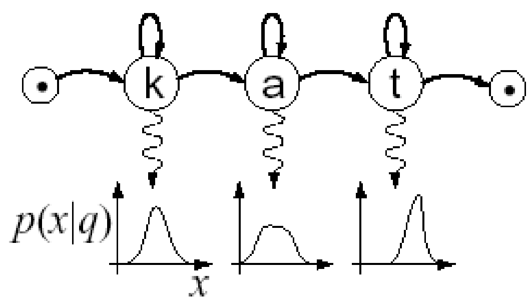
- states q^i

- transition probabilities a_{ij}



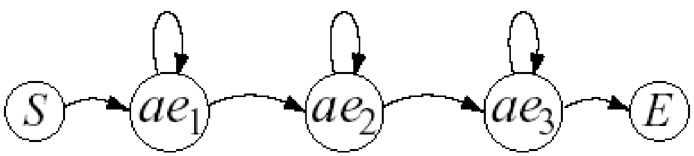
	k	a	t	•
•	1.0	0.0	0.0	0.0
k	0.9	0.1	0.0	0.0
a	0.0	0.9	0.1	0.0
t	0.0	0.0	0.9	0.1

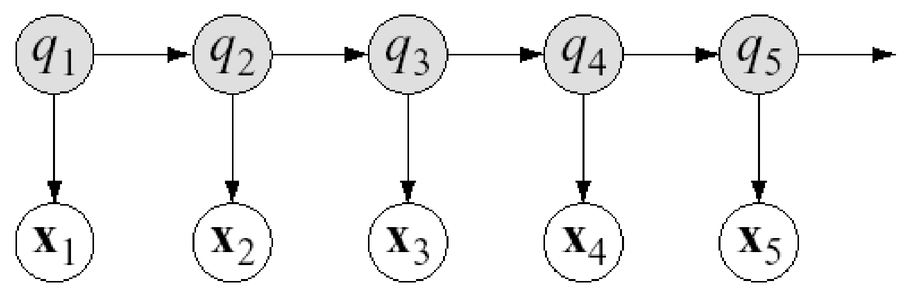
- emission distributions $b_i(x)$



Markov models for speech

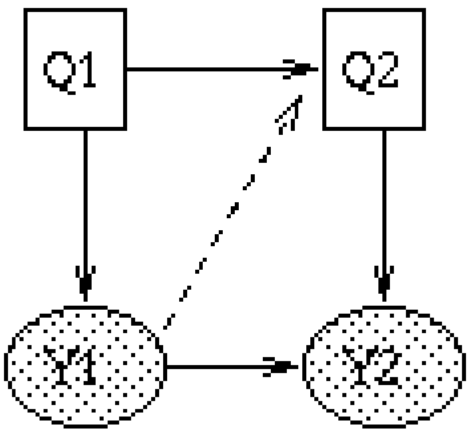
- **Speech models** M_j
  - typ. left-to-right HMMs (sequence constraint)
  - observation & evolution are conditionally independent of rest given (hidden) state q_n



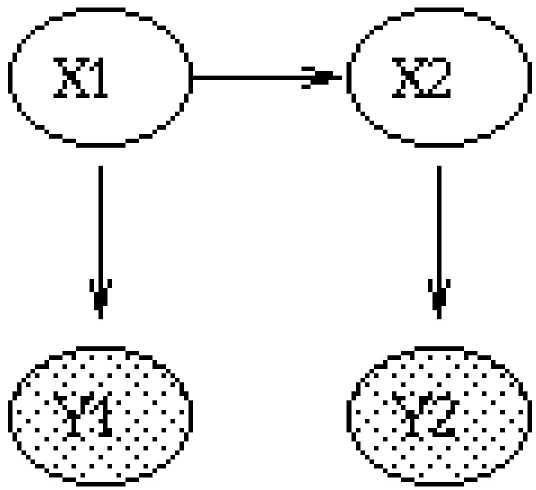


  - *self-loops* for time dilation

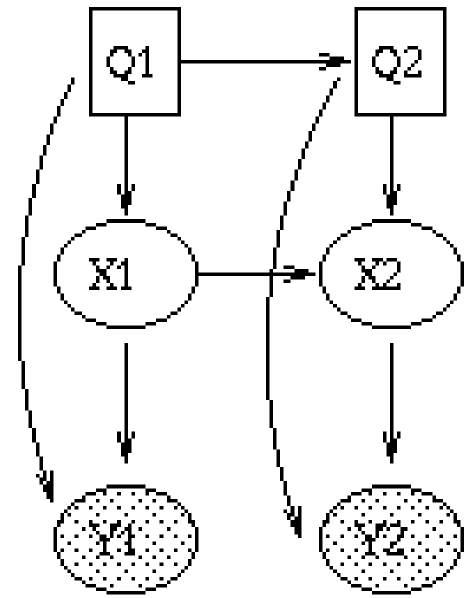
Types of Sequential State Models



Switching AR model



Kalman filter model



Switching Kalman filter

Y: Observed
X: State
Q: Discrete State (e.g. Decision)

Transition matrix where initial state matters

$$\begin{array}{c} \\ S_1 \\ S_2 \\ S_3 \\ S_4 \end{array} \begin{array}{c} \begin{array}{cccc} S_1 & S_2 & S_3 & S_4 \end{array} \\ \begin{bmatrix} p_{11} & p_{12} & 0 & 0 \\ p_{21} & p_{22} & 0 & 0 \\ 0 & 0 & p_{33} & p_{34} \\ 0 & 0 & p_{43} & p_{44} \end{bmatrix} \end{array}.$$

State update

$$[p(F_1) \quad p(U_1)] \begin{bmatrix} p(F_2|F_1) & p(U_2|F_1) \\ p(F_2|U_1) & p(U_2|U_1) \end{bmatrix} = [p(F_2) \quad p(U_2)].$$

Guessing Games

"Belief in the law of Small Numbers"

Guess the next:

- HTHHHH
- HHHTTT
- HTHTHT

Rule 1: Estimating the frequency

Rule 2: Using serial prediction

Rule 3: Estimating the likelihood

All-or-None Learning Models

S_1 The preinsight state

S_2 The postinsight state

X_i Animal's current state

$$X_i = S_1 \quad \text{for} \quad i < k,$$

$$X_i = S_2 \quad \text{for} \quad i \geq k.$$

All or None Learning Models

Observer generates a set of different sorting hypotheses
Color, Mixture of Suits, Face vs Number, etc.
Observer tries a hypothesis and told if sort is correct

S1 = Incorrect Hyp, Told Incorrect
S2 = Correct, Told Correct

Random Hypothesis Selection

$$\text{Model 1:}\quad \begin{array}{c} \\ S_1 \\ S_2 \end{array} \begin{array}{c} \begin{array}{cc} S_1 & S_2 \end{array} \\ \begin{bmatrix} (m-1)/m & 1/m \\ 0 & 1 \end{bmatrix} \end{array}.$$

Stationary

Process of Elimination

$$\text{Model 2:}\quad \begin{array}{c} \\ S_1 \\ S_2 \end{array} \begin{array}{c} \begin{array}{cc} S_1 & S_2 \end{array} \\ \begin{bmatrix} (m-n)/(m-n+1) & 1/(m-n+1) \\ 0 & 1 \end{bmatrix} \end{array}$$

Non-Stationary

Response Models

Card Sorting: subject has to determine sorting rule, seeing two cards lying face up.

Sort by color R-B

Observer's X = Card

S1 = { 2H,2D, …., AceH,AceD}

S2 = { 2S,2C, …., AceS,AceC}

Sequential Games

C = Cooperate
D = Defect

		Player 2 C	Player 2 D
Player 1	C	+0.25, +0.25	−1.00, +1.00
Player 1	D	+1.00, −1.00	−0.25, −0.25

Male-Female Pair-off Differences (Rapoport & Chammah)

TABLE 8.1 PROPORTION OF COOPERATIVE AND DEFECTING STRATEGIES OF DIFFERENT TYPES OF PAIRS

	$\hat{p}(CC)$	$\hat{p}(CD)$	$\hat{p}(DC)$	$\hat{p}(DD)$	$\hat{p}(C)$
MM	.51	.08	.09	.32	.59
WM	.40	.10	.10	.41	.49
WW	.23	.11	.11	.55	.34

Additional Analysis

$p(C_n|C_{n-1}, C'_{n-1})$: The probability an individual makes a cooperative choice on trial n given both he and his opponent made a cooperative choice on trial $n-1$

$p(C_n|C_{n-1}, D'_{n-1})$: The probability an individual makes a cooperative choice on trial n given he made a cooperative choice on trial $n-1$ and his opponent made a defecting choice

$p(C_n|D_{n-1}, C'_{n-1})$: The probability an individual makes a cooperative choice on trial n given he made a defecting choice on trial $n-1$ and his opponent made a cooperative choice

$p(C_n|D_{n-1}, D'_{n-1})$: The probability an individual makes a cooperative choice on trial n given both he and his opponent made a defecting choice on trial $n-1$

Data

	$\hat{p}(C_n \mid C_{n-1}, C'_{n-1})$	$\hat{p}(C_n \mid C_{n-1}, D'_{n-1})$	$\hat{p}(C_n \mid D_{n-1}, C'_{n-1})$	$\hat{p}(C_n \mid D_{n-1}, D'_{n-1})$
MM	0.85	0.40	0.38	0.20
WM	0.79	0.42	0.31	0.22
WW	0.75	0.37	0.26	0.15

Fig. 8.4 Estimated probability of choosing the cooperative strategy conditioned on the preceding choices.

Belief in Sequential Response Dependence

The Cold Facts About the "Hot Hand" in Basketball

Amos Tversky and Thomas Gilovich

Do Players hit shots in streaks?

Reasonability: Recalibration

Beliefs

agreement: 91% of the fans believed that a player has "a better chance of making a shot after having just *made* his last two or three shots than he does after having just *missed* his last two or three shots"; 68% of the fans expressed essentially the same belief for free throws, claiming that a player has "a better chance of making his second shot after *making* his first shot than after *missing* his first shot"; 96% of the fans thought that "after having made a series of shots in a row . . . players tend to take more shots than they normally would"; 84% of the fans believed that "it is important to pass the ball to someone who has just made several (two, three, or four) shots in a row."

Conditional Prob for Players

TABLE 1

Probability of Making a Shot Conditioned on the Outcome of Previous Shots for Nine Members of the Philadelphia 76ers

Player	*P*(hit/3 misses)	*P*(hit/2 misses)	*P*(hit/1 miss)	*P*(hit)	*P*(hit/1 hit)	*P*(hit/2 hits)	*P*(hit/3 hits)	Serial correlation *r*
Clint Richardson	.50 (12)	.47 (32)	.56 (101)	.50 (248)	.49 (105)	.50 (46)	.48 (21)	−.020
Julius Erving	.52 (90)	.51 (191)	.51 (408)	.52 (884)	.53 (428)	.52 (211)	.48 (97)	.016
Lionel Hollins	.50 (40)	.49 (92)	.46 (200)	.46 (419)	.46 (171)	.46 (65)	.32 (25)	−.004
Maurice Cheeks	.77 (13)	.60 (38)	.60 (126)	.56 (339)	.55 (166)	.54 (76)	.59 (32)	−.038
Caldwell Jones	.50 (20)	.48 (48)	.47 (117)	.47 (272)	.45 (108)	.43 (37)	.27 (11)	−.016
Andrew Toney	.52 (33)	.53 (90)	.51 (216)	.46 (451)	.43 (190)	.40 (77)	.34 (29)	−.083
Bobby Jones	.61 (23)	.58 (66)	.58 (179)	.54 (433)	.53 (207)	.47 (96)	.53 (36)	−.049
Steve Mix	.70 (20)	.56 (54)	.52 (147)	.52 (351)	.51 (163)	.48 (77)	.36 (33)	.015
Daryl Dawkins	.88 (8)	.73 (33)	.71 (136)	.62 (403)	.57 (222)	.58 (111)	.51 (55)	−.142**
Weighted means	.56	.53	.54	.52	.51	.50	.46	−.039

Note. Since the first shot of each game cannot be conditioned, the parenthetical values in columns 4 and 6 do not sum to the parenthetical value in column 5. The number of shots upon which each probability is based is given in parentheses.

* $p < .05$.

** $p < .01$.

Philadelphia 76ers 1980-81 season

Runs Test

TABLE 2
Runs Test—Philadelphia 76ers

Players	Hits	Misses	Number of runs	Expected number of runs	Z
Clint Richardson	124	124	128	125.0	−0.38
Julius Erving	459	425	431	442.4	0.76
Lionel Hollins	194	225	203	209.4	0.62
Maurice Cheeks	189	150	172	168.3	−0.41
Caldwell Jones	129	143	134	136.6	0.32
Andrew Toney	208	243	245	225.1	−1.88
Bobby Jones	233	200	227	216.2	−1.04
Steve Mix	181	170	176	176.3	0.04
Daryl Dawkins	250	153	220	190.8	−3.09**
M =	218.6	203.7	215.1	210.0	−0.56

* $p < .05$.
** $p < .01$.

Runs: consecutive hits or misses
X000XX0 => 4 “runs”

Free Throw Data

TABLE 3
Probability of Making a Second Free Throw Conditioned on the Outcome of the First Free Throw for Nine Members of the Boston Celtics during the 1980–1981 and 1981–1982 Seasons

Player	$P(H_2/M_1)$	$P(H_2/H_1)$	Serial correlation r
Larry Bird	.91 (53)	.88 (285)	−.032
Cedric Maxwell	.76 (128)	.81 (302)	.061
Robert Parish	.72 (105)	.77 (213)	.056
Nate Archibald	.82 (76)	.83 (245)	.014
Chris Ford	.77 (22)	.71 (51)	−.069
Kevin McHale	.59 (49)	.73 (128)	.130
M. L. Carr	.81 (26)	.68 (57)	−.128
Rick Robey	.61 (80)	.59 (91)	−.019
Gerald Henderson	.78 (37)	.76 (101)	−.022

Note. The number of shots upon which each probability is based is given in parentheses.

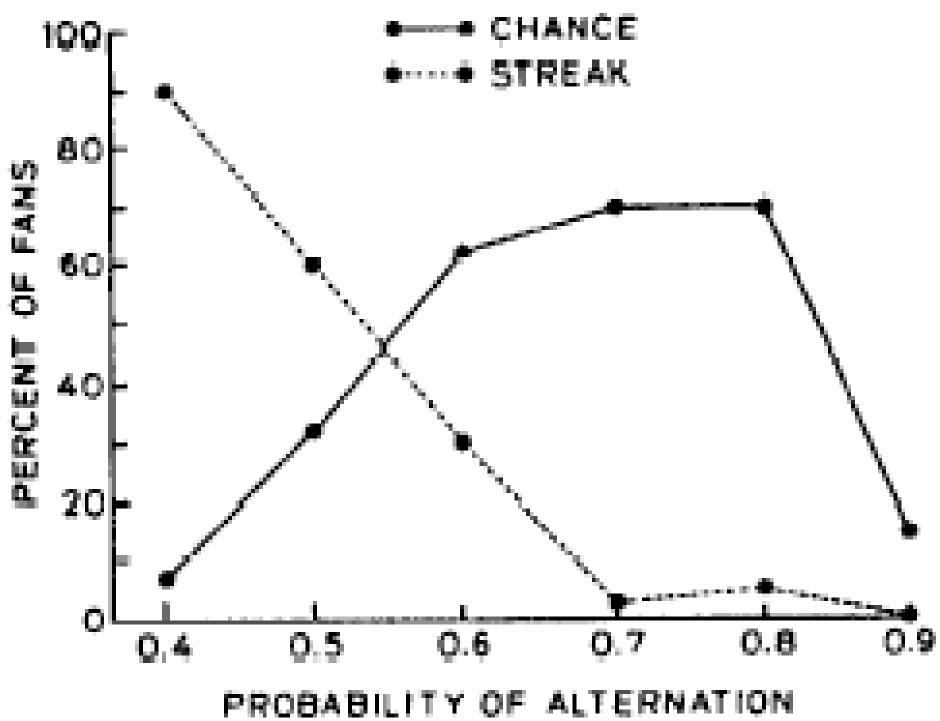


FIG. 1. Percentage of basketball fans classifying sequences of hits and misses as examples of streak shooting or chance shooting, as a function of the probability of alternation within the sequences.

Probability of Alternation
P(H|Miss)

Erroneous “Law of small numbers”

Kahneman and Tversky (1982a, p. 44) illustrate how people expect close to the same probability distribution of types in small groups as they do in large groups, asking a group of undergraduates the following question:

> A certain town is served by two hospitals. In the larger hospital about 45 babies are born each day, and in the smaller hospital about 15 babies are born each day. As you know, about 50 percent of all babies are boys. However, the exact percentage varies from day to day. Sometimes it may be higher than 50 percent, sometimes lower. For a period of 1 year, each hospital recorded the days on which more than 60 percent of the babies born were boys. Which hospital do you think recorded more such days?

22% Larger hospital => more days over 60
56% The same
22% Smaller hospital => more days over 60

Producing Random sequences

Rapoport and Budescu (1992, 1997, 1994)
asked subjects

- "simulate the random outcome of tossing an unbiased coin 150 times in succession,"
- "imagine a sequence of 150 draws with replacement from a well-shuffled deck, including five red and five black cards, and then call aloud the sequence of these binary draws.

Results

$\Pr(A \mid B)$	58.5%
$\Pr(A \mid AB)$	46.0%
$\Pr(A \mid AAB)$	38.0%
$\Pr(A \mid AAA...)$	29.8%

Can we produce random sequences in games?

Walker and Wooders (1999)
final and semi-final matches at Wimbledon

"Our tests indicate that the tennis players are not quite playing randomly: they switch their serves from left to right and vice versa somewhat too often to be consistent with random play. This is consistent with extensive experimental research in psychology and economics which indicates that people who are attempting to behave truly randomly tend to ''switch too often.'' "

Perception of Randomness

Randomness and Coincidences:
Reconciling Intuition and Probability Theory
Thomas L. Griffiths & Joshua B. Tenenbaum

Randomness as a rational inference - Belief that most sequences have non-random causes

$$\log \frac{P(\text{random}|x)}{P(\text{regular}|x)},$$

Zenith Radio “Psychic transmissions”

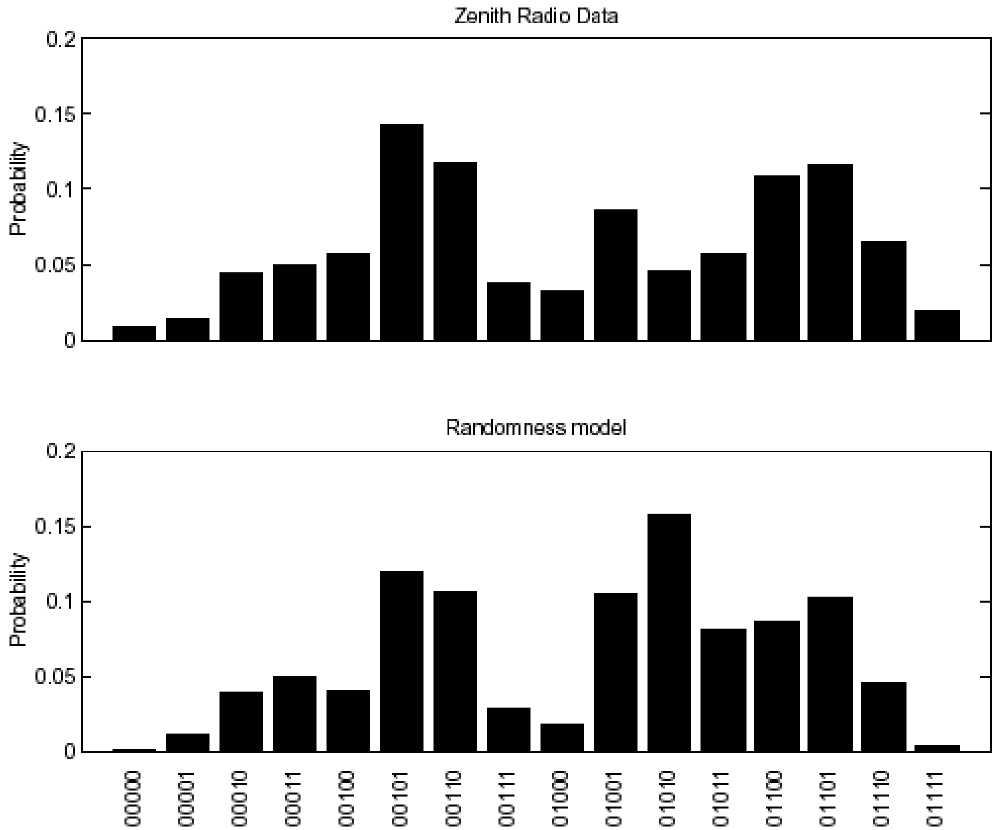


Figure 1: The upper panel shows the original Zenith radio data, representing the responses of 20,099 participants, from Goodfellow (1938). The lower panel shows the predictions of the randomness model. Sequences are collapsed over the initial choice, represented by 0.

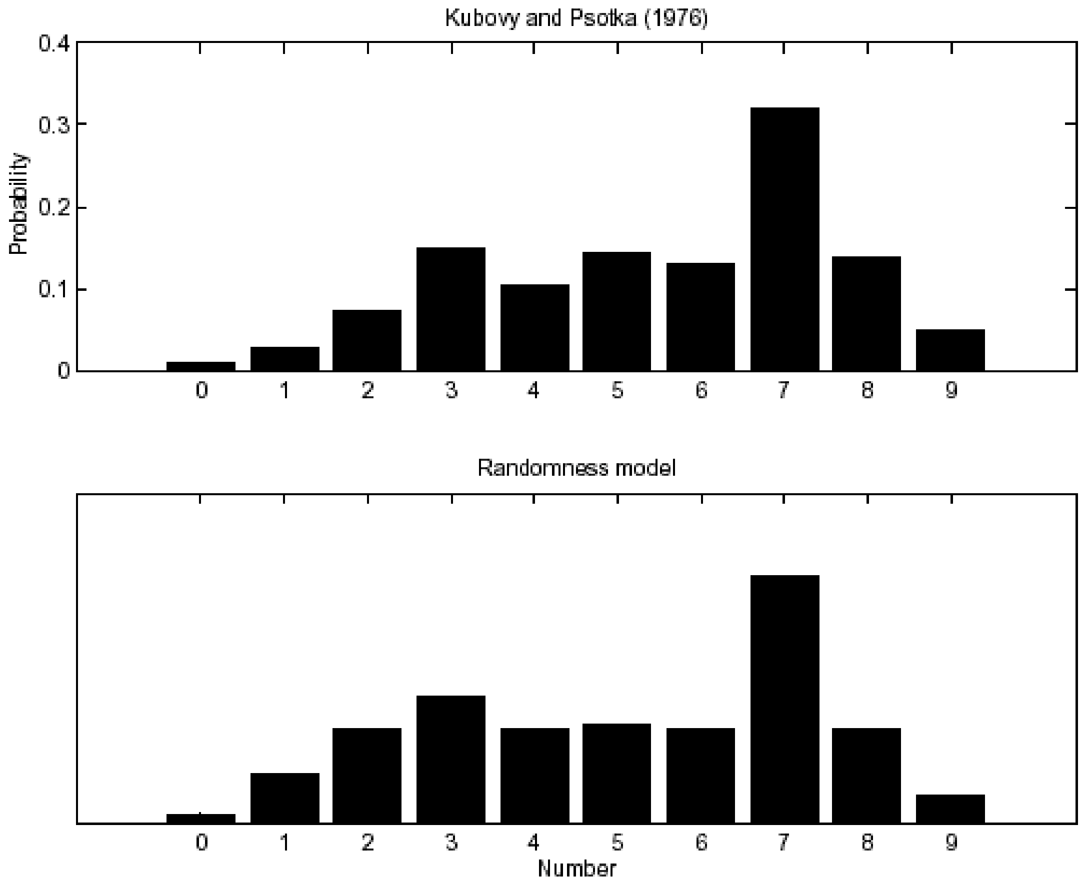
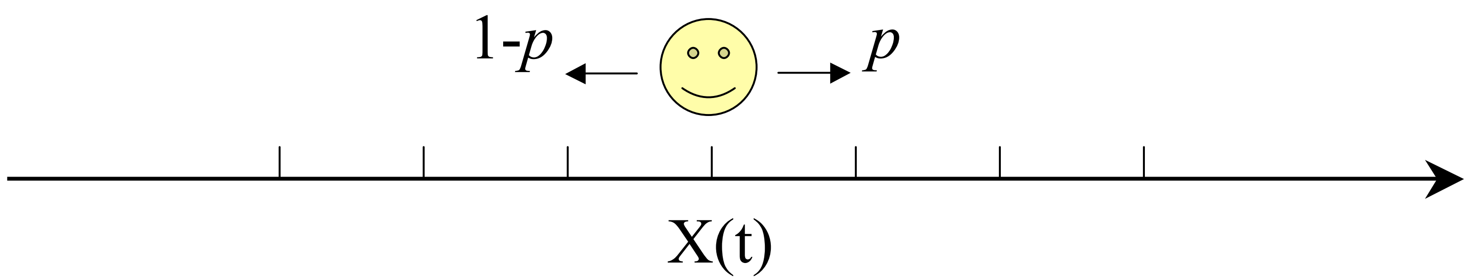


Figure 2: The upper panel shows number production data from Kubovy and Psotka (1976), taken from 1,770 participants choosing numbers between 0 and 9. The lower panel shows the transformed predictions of the randomness model.

1-D Random Walk



- Time is slotted
- The walker flips a coin every time slot to decide which way to go
-

Transition Probability

- Probability to jump from state *i* to state *j*
- Assume **stationary**: independent of time
- Transition probability matrix:

$$P = (p_{ij})$$

- Two state MC:

$$P = \begin{pmatrix} 1-p & p \\ q & 1-q \end{pmatrix}$$

Stationary Distribution

Define

Then $\pi_{k+1} = \pi_k P$ (π is a row vector)

Stationary Distribution:

if the limit exists.

$$\pi_k(i) = \text{Pr}\{X_k = i\}$$

$$\pi = \lim_{k\to\infty} \pi_k$$

If π exists, we can solve it by

$$\pi = \pi P, \quad \sum_i \pi(i) = 1$$

Balance Equations

- These are called **balance equations**
  - Transitions in and out of state *i* are balanced

In General

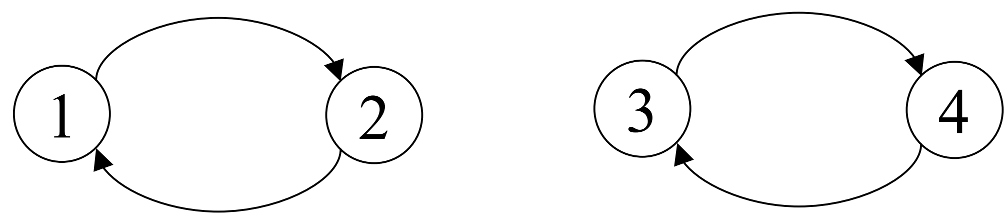
- If we partition all the states into two sets, then transitions between the two sets must be “balanced”.
  - Equivalent to a bi-section in the state transition graph
  - This can be easily derived from the Balance Equations

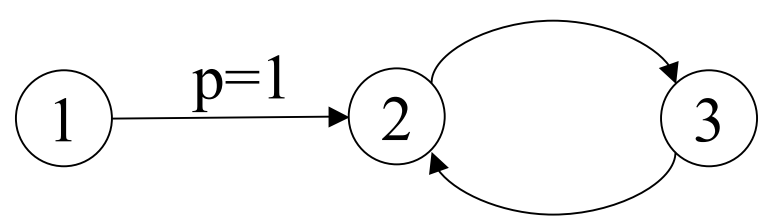
Conditions for π to Exist (I)

- Definitions:
  - State j is **reachable** by state i if
  - State i and j **commute** if they are reachable by each other
  - The Markov chain is **irreducible** if all states commute

Conditions for π to Exist (I) (cont'd)

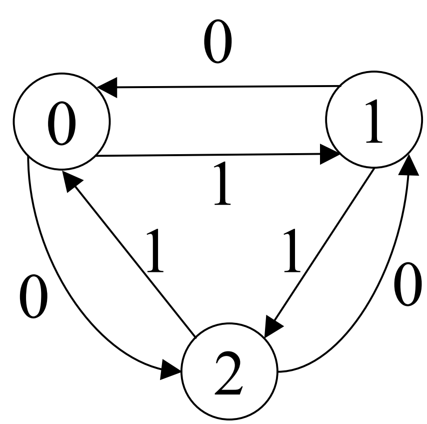
- Condition: The Markov chain is **irreducible**
- Counter-examples:





Conditions for π to Exist (II)

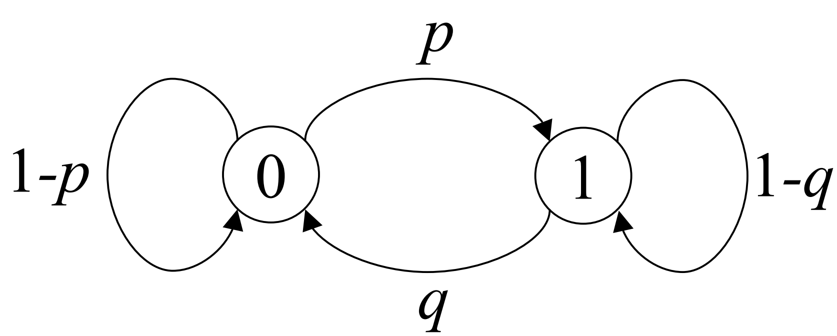
- The Markov chain is **aperiodic**:
- Counter-example:



Conditions for π to Exist (III)

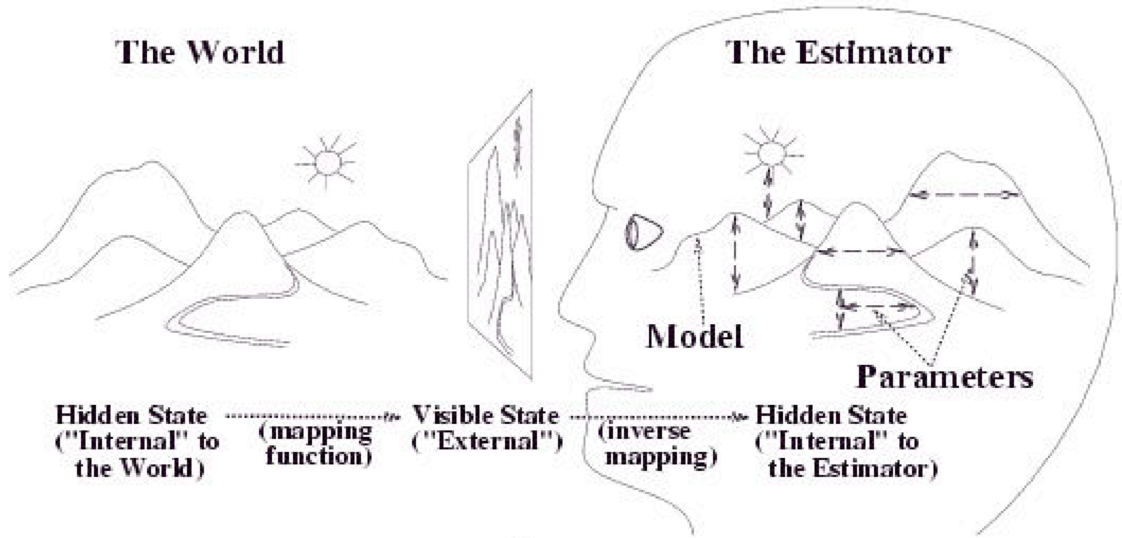
- The Markov chain is **positive recurrent**:
  - State *i* is **recurrent** if
  - Otherwise **transient**
  - If recurrent
    - State *i* is **positive recurrent** if $E(T_i)<1$, where T_i is time between visits to state *i*
    - Otherwise **null recurrent**

Solving for π



$$P = \begin{pmatrix} 1-p & p \\ q & 1-q \end{pmatrix}$$

Sequential Models of Perception



(a)