Particle Filter Failures

References

King and Forsyth, "How Does CONDENSATION Behave with a Finite Number of Samples?" ECCV 2000, 695-709.

Karlin and Taylor, A First Course in Stochastic Processes, 2nd edition, Academic Press, 1975.

Particle Filter Failures

Summary

Condensation/SIR is aymptotically correct as the number of samples tends towards infinity. However, as a practical issue, it has to be run with a finite number of samples.

Iterations of Condensation form a Markov chain whose state space is quantized representations of a density.

This Markov chain has some undesirable properties

- high variance - different runs can lead to very different answers
- low apparent variance within each individual run (appears stable)
- state can collapse to single peak in time roughly linear in number of samples
- tracker may appear to follow peaks in the posterior even in the absence of any meaningful measurements.

These properties generally known as “sample impoverishment”

Stationary Analysis

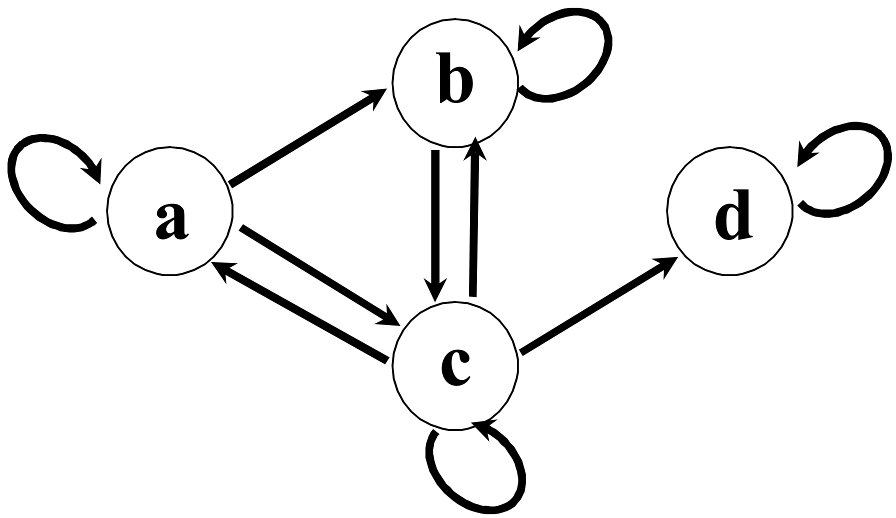
For simplicity, we focus on tracking problems with stationary distributions (posterior should be the same at any time step).

[because it is hard to really focus on what is going on when the posterior modes are deterministically moving around. Any movement of modes in our analysis will be due to behavior of the particle filter]

A Little Stochastic Process Theory

Markov Chain:

- A sequence of random variables $X_1, X_2, X_3, \ldots$
- Each variable is a distribution over a set of states (a,b,c...)
- Transition probability of going to next state only depends on the current state. e.g. $P(X_{n+1} = a \mid X_n = b)$

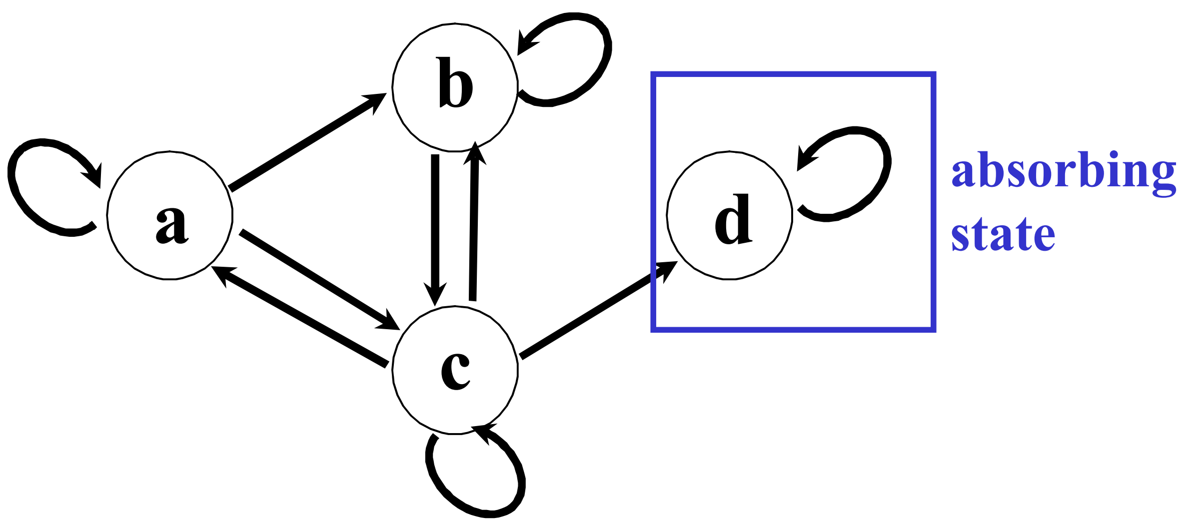


A Little Stochastic Process Theory

Absorbing states are those from which you can never leave.

If there is a nonzero probability of reaching an absorbing state, you eventually will

-Murphy's law applied to stochastic processes!



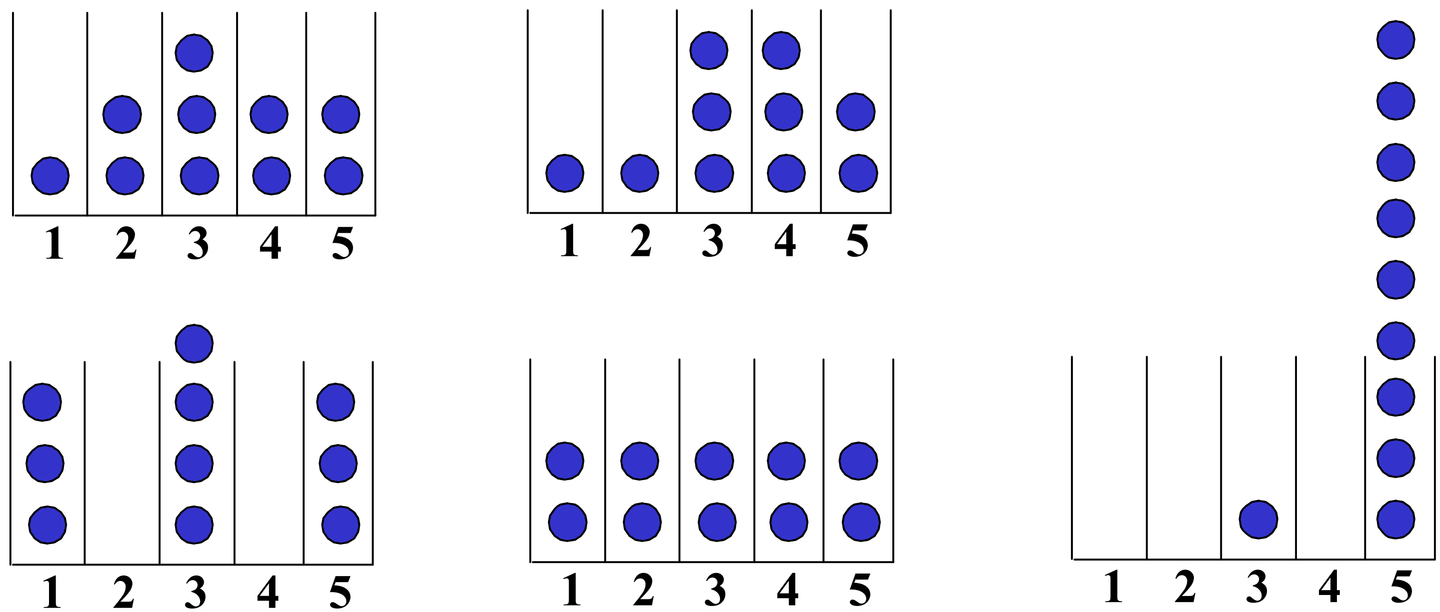
Jump up a Level of Abstraction

- Consider each state to be a configuration of particles.
- Carefully consider how particle filtering moves from one state (configuration) to another
- Define a probabilistic model for those transitions
- Study the resulting Markov Chain stochastic process

Illustration: PMF State Space

Consider 10 particles representing a probability mass function (pmf) over 5 locations.

Examples of pmf state space

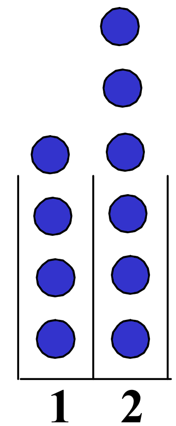


Even Simpler PMF State Space

Consider 10 particles representing a probability mass function over 2 locations.

PMF state space:

**{(0,10)(1,9)(2,8)(3,7)(4,6)(5,5)
(6,4)(7,3)(8,2)(9,1)(10,0)}**



(4,6)

We will now instantiate a particular two-state filtering model that we can analyze in closed-form, and explore the Markov chain process (on the PMF state space above) that describes how particle filtering performs on that process.

Discrete, Stationary, No Noise

Assume a stationary process model with no-noise

process model: $X_{k+1} = \not{F}\, X_k + \not{v}_k$

I **0**

Identity **no noise**

process model: $X_{k+1} = X_k$

Perfect Two-State Ambiguity

Let our two filtering states be {a,b}.

We define both prior distribution and observation model to be ambiguous (equal belief in a and b).

$$P(X_0) = \begin{cases} .5 & X_0 = a \\ .5 & X_0 = b \end{cases} \qquad P(Z|X_k) = \begin{cases} .5 & X_0 = a \\ .5 & X_0 = b \end{cases}$$

from process model:

$P(X_{k+1} \mid X_k) =$

	a	b
a	1	0
b	0	1

Recall: Recursive Filtering

Prediction:

predicted current state — **state transition** — **previous estimated state**

$$p(\mathbf{x}_k|\mathbf{z}_{1:k-1}) = \int p(\mathbf{x}_k|\mathbf{x}_{k-1})p(\mathbf{x}_{k-1}|\mathbf{z}_{1:k-1})\, d\mathbf{x}_{k-1}.$$

Update:

estimated current state — **measurement** — **predicted current state**

$$p(\mathbf{x}_k|\mathbf{z}_{1:k}) = \frac{p(\mathbf{z}_k|\mathbf{x}_k)p(\mathbf{x}_k|\mathbf{z}_{1:k-1})}{p(\mathbf{z}_k|\mathbf{z}_{1:k-1})}$$

normalization term

$$p(\mathbf{z}_k|\mathbf{z}_{1:k-1}) = \int p(\mathbf{z}_k|\mathbf{x}_k)p(\mathbf{x}_k|\mathbf{z}_{1:k-1})\, d\mathbf{x}_k$$

These are exact propagation equations.

Analytic Filter Analysis

Predict

$$p(\mathbf{x}_k|\mathbf{z}_{1:k-1}) = \int p(\mathbf{x}_k|\mathbf{x}_{k-1})p(\mathbf{x}_{k-1}|\mathbf{z}_{1:k-1})\, d\mathbf{x}_{k-1}.$$

$$P(X_1=a|z_{1:0}) = \overset{\mathbf{1}}{P(X_1=a|X_0=a)}\overset{\mathbf{.5}}{P(X_0=a)} + \overset{\mathbf{0}}{P(X_1=a|X_0=b)}\overset{\mathbf{.5}}{P(X_0=b)} \quad \mathbf{= .5}$$

$$P(X_1=b|z_{1:0}) = \overset{\mathbf{0}}{P(X_1=b|X_0=a)}\overset{\mathbf{.5}}{P(X_0=a)} + \overset{\mathbf{1}}{P(X_1=b|X_0=b)}\overset{\mathbf{.5}}{P(X_0=b)} \quad \mathbf{= .5}$$

Update

$$p(\mathbf{x}_k|\mathbf{z}_{1:k}) = \frac{p(\mathbf{z}_k|\mathbf{x}_k)p(\mathbf{x}_k|\mathbf{z}_{1:k-1})}{\int p(\mathbf{z}_k|\mathbf{x}_k)p(\mathbf{x}_k|\mathbf{z}_{1:k-1})\, d\mathbf{x}_k}$$

$$P(X_1=a|z_{1:1}) \propto \overset{\mathbf{.5}}{P(z_1|X_1=a)}\overset{\mathbf{.5}}{P(X_1=a|z_{1:0})} \quad \mathbf{= .25/(.25+.25) = .5}$$

$$P(X_1=b|z_{1:1}) \propto \overset{\mathbf{.5}}{P(z_1|X_1=b)}\overset{\mathbf{.5}}{P(X_1=b|z_{1:0})} \quad \mathbf{= .25/(.25+.25) = .5}$$

Analytic Filter Analysis

Therefore, for all k, the posterior distribution is

$$P(X_k \mid z_{1:k}) = \begin{cases} .5 & X_k = a \\ .5 & X_k = b \end{cases}$$

which agrees with our intuition in regards to the stationarity and ambiguity of our two-state model.

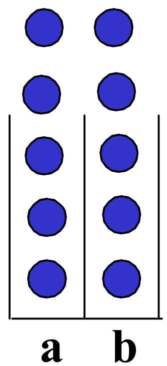
Now let's see how a particle filter behaves...

Particle Filter

Consider 10 particles representing a probability mass function over our 2 locations {a,b}.

In accordance with our ambiguous prior, we will initialize with 5 particles in each location

$P(X_0) =$



Condensation (SIR) Particle Filter

1) Select N new samples with replacement, according to the sample weights **(equal weights in this case)**
2) Apply process model to each sample (deterministic motion + noise) **(no-op in this case)**
3) For each new position, set weight of particle in accordance to observation probability

   (all weights become .5 in this case)
4) Normalize weights so they sum to one

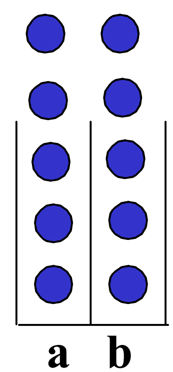
   (weights are still equal)

Condensation as Markov Chain
(Key Step)

Recall that 10 particles representing a probability mass function over 2 locations can be thought of as having a state space with 11 elements:

{(0,10)(1,9)(2,8)(3,7)(4,6)(5,5)(6,4)(7,3)(8,2)(9,1)(10,0)}

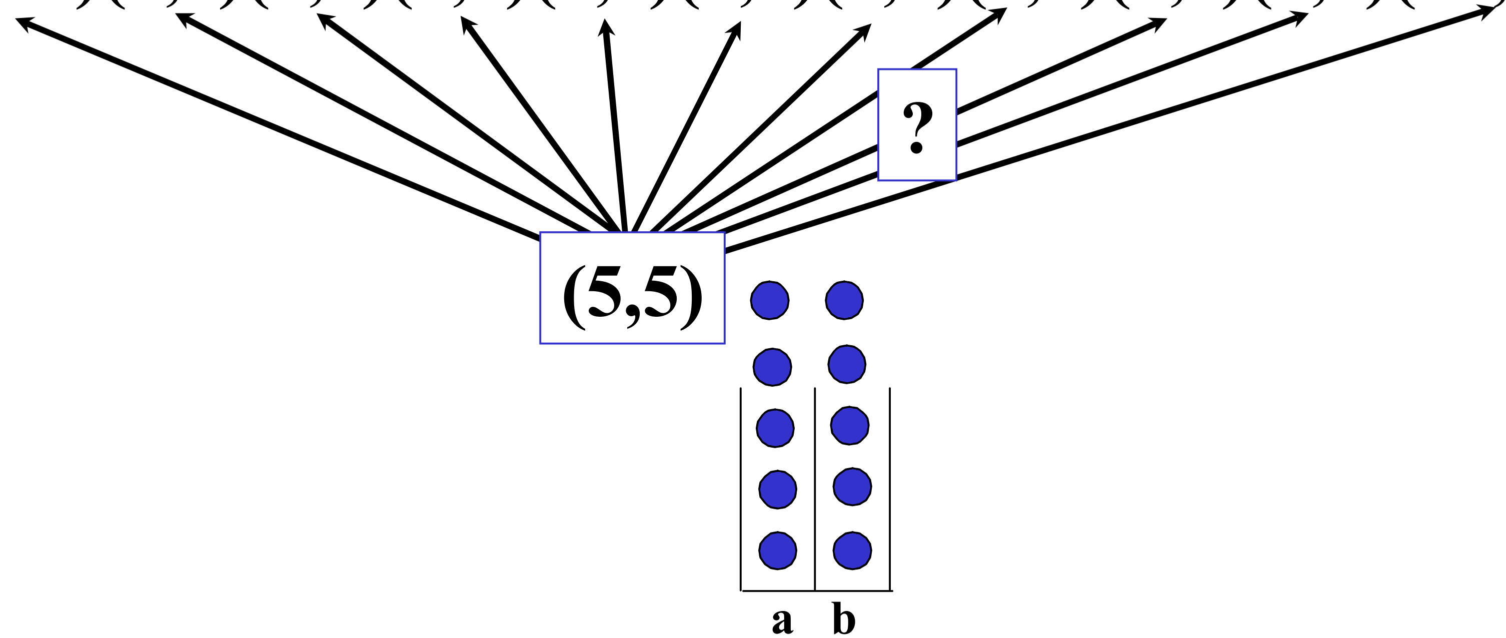
(5,5)



Condensation as Markov Chain
(Key Step)

We want to characterize the probability that the particle filter procedure will transition from the current configuration to a new configuration:

{(0,10)(1,9)(2,8)(3,7)(4,6)(5,5)(6,4)(7,3)(8,2)(9,1)(10,0)}

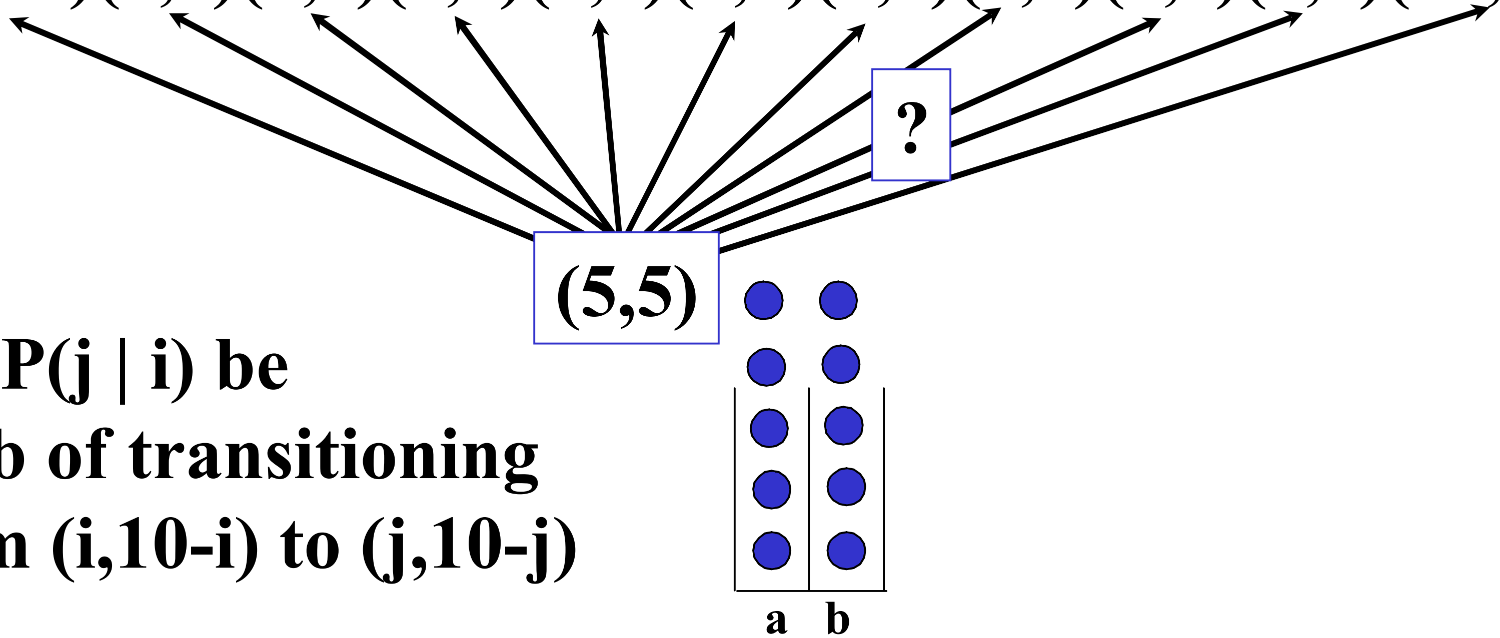


Robert Collins
CSE598C

Condensation as Markov Chain
(Key Step)

We want to characterize the probability that the particle filter procedure will transition from the current configuration to a new configuration:

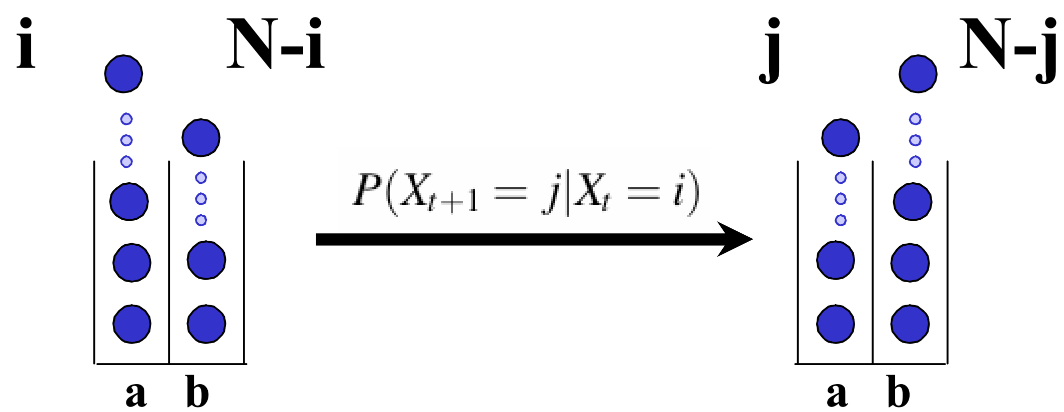
{(0,10)(1,9)(2,8)(3,7)(4,6)(5,5)(6,4)(7,3)(8,2)(9,1)(10,0)}



Let P(j | i) be prob of transitioning from (i,10-i) to (j,10-j)

Robert Collins
CSE598C

Transition Probability



$$P(X_{t+1} = j | X_t = i) = \binom{N}{j} \left(\frac{i}{N}\right)^j \left(\frac{N-i}{N}\right)^{N-j}$$

number of ways of selecting j points out of N times (since we don't care about the order in which we select the j points)

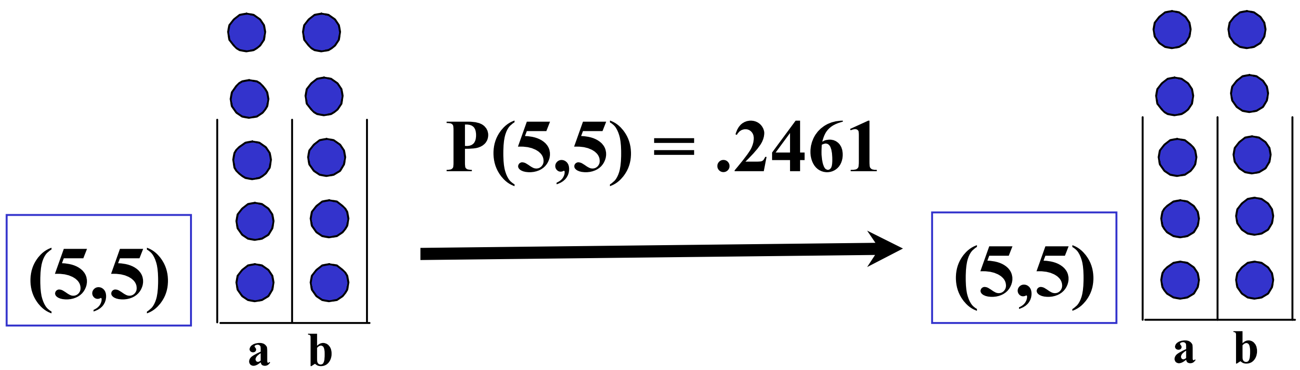
probability of choosing j points from first bucket (with prob i/N)

probability of choosing N-j points from second bucket (with prob (N-i)/N)

Example

$$P(X_{t+1} = j | X_t = i) = \binom{N}{j} \left(\frac{i}{N}\right)^j \left(\frac{N-i}{N}\right)^{N-j}$$

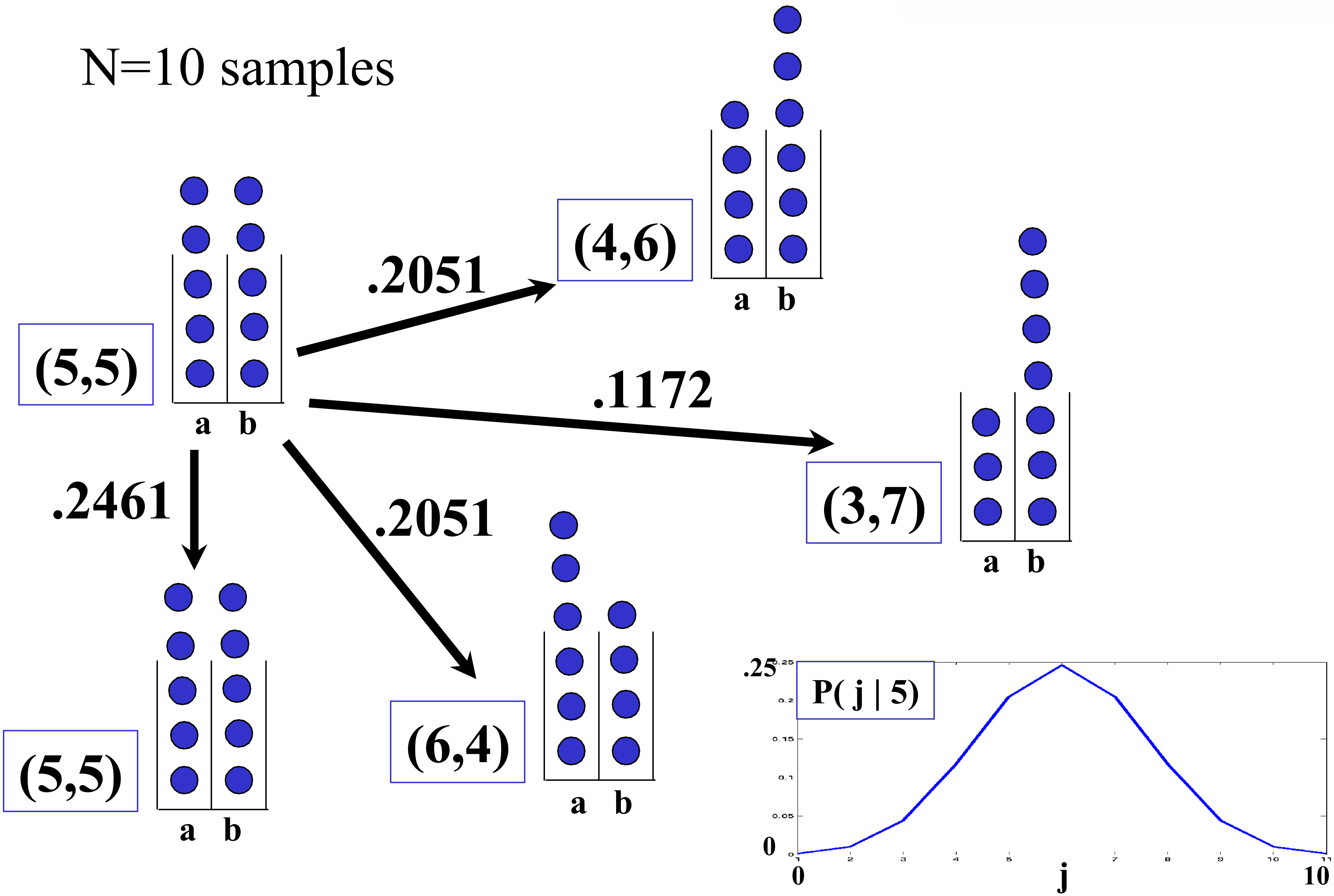
N=10 samples



Example

$$\binom{N}{j}\left(\frac{i}{N}\right)^{j}\left(\frac{N-i}{N}\right)^{N-j}$$

N=10 samples

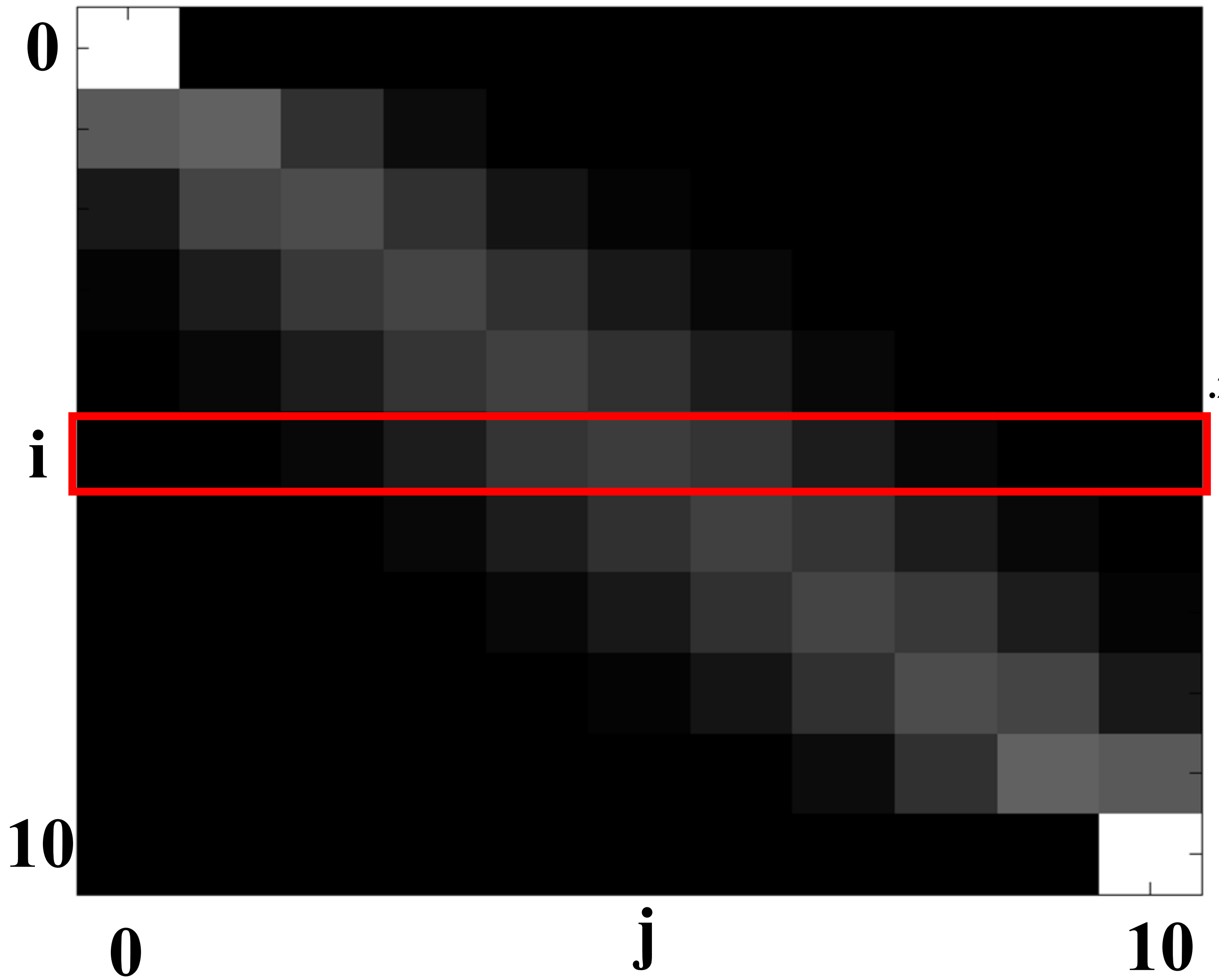


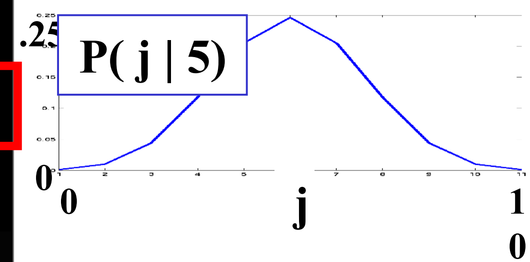
Robert Collins
CSE598C

Full Transition Table

$$\binom{N}{j}\left(\frac{i}{N}\right)^{j}\left(\frac{N-i}{N}\right)^{N-j}$$

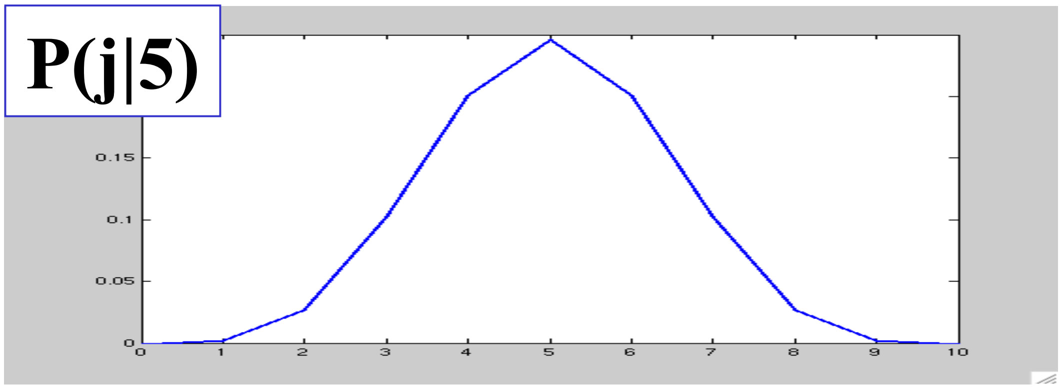
P(j | i)



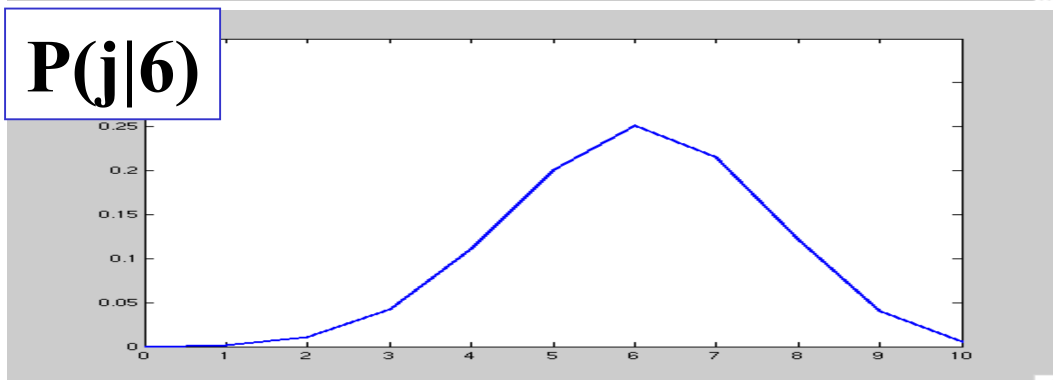


Robert Collins
CSE598C

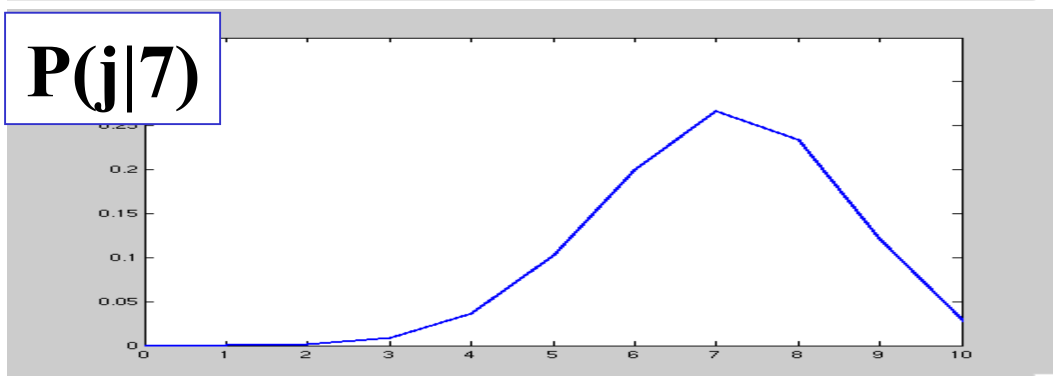
The Crux of the Problem



from (5,5), there is a good chance we will jump to away from (5,5), say to (6,4)



once we do that, we are no longer sampling from the transition distribution at (5,5), but from the one at (6,4). But this is biased off center from (5,5)



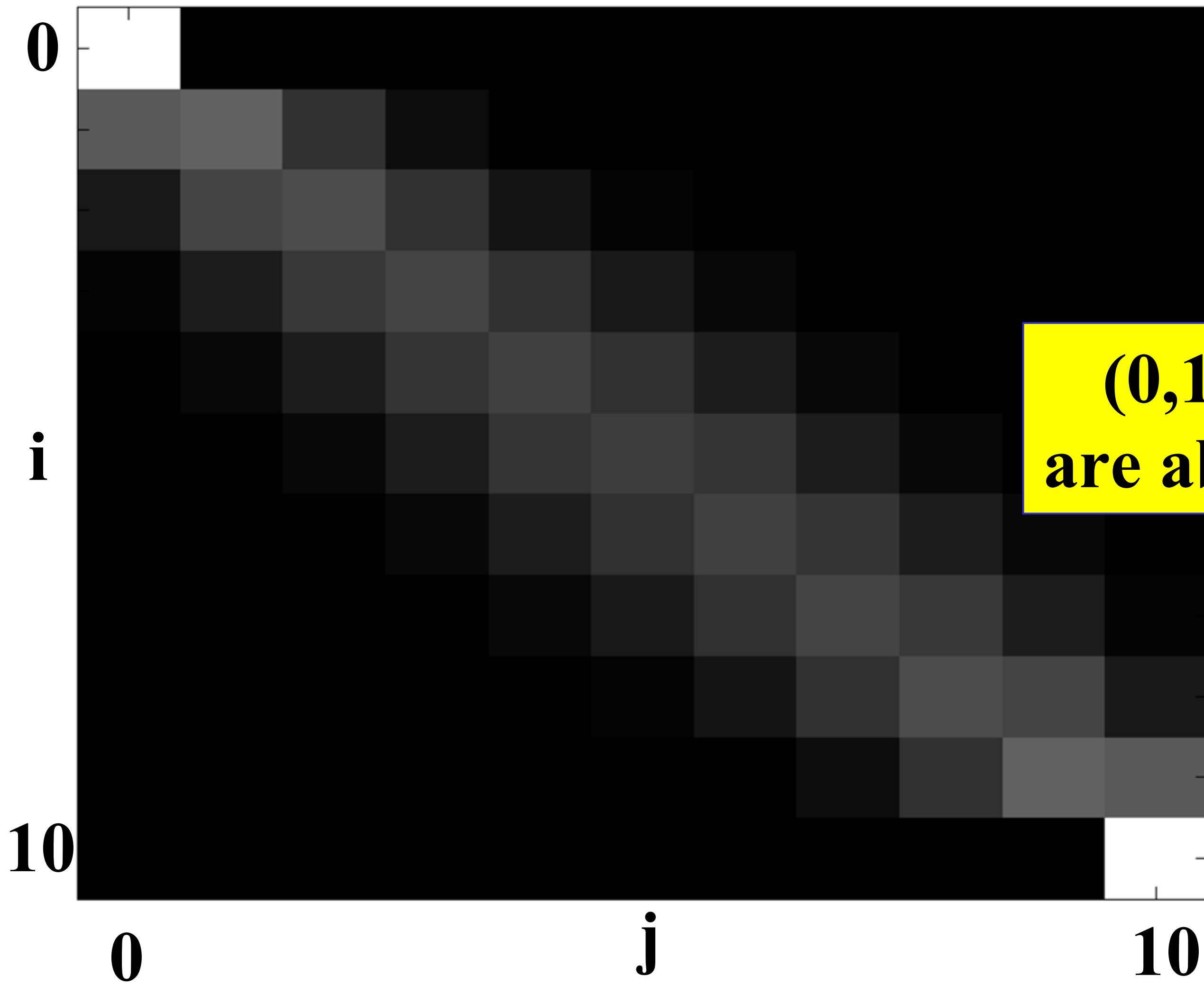
and so on. The behavior will be similar to that of a random walk.

Robert Collins
CSE598C

Another Problem

$$\binom{N}{j}\left(\frac{i}{N}\right)^{j}\left(\frac{N-i}{N}\right)^{N-j}$$

P(j | i)



P(0|0) = 1

(0,10) and (10,0) are absorbing states!

P(10|10) = 1

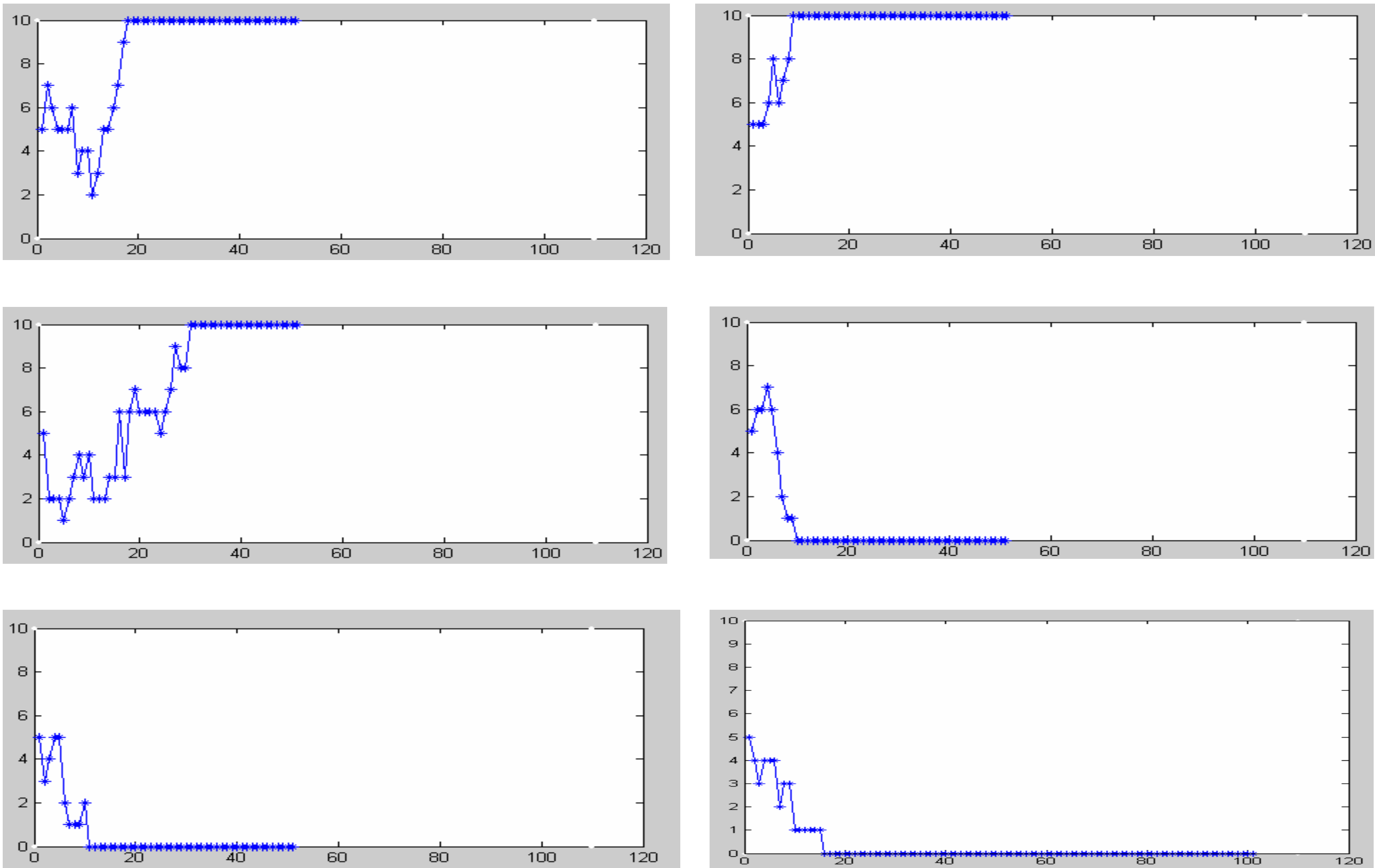
Observations

- The Markov chain has two absorbing states (0,10) and (10,0)
- Once the chain gets into either of these two states, it can never get out (all the particles have collapsed into a single bucket)
- There is a nonzero probability of getting into either absorbing state, starting from (5,5)

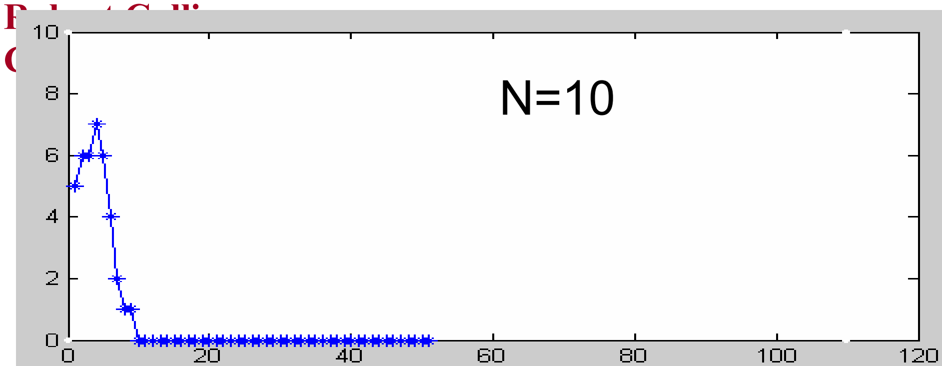
These are the seeds of our destruction!

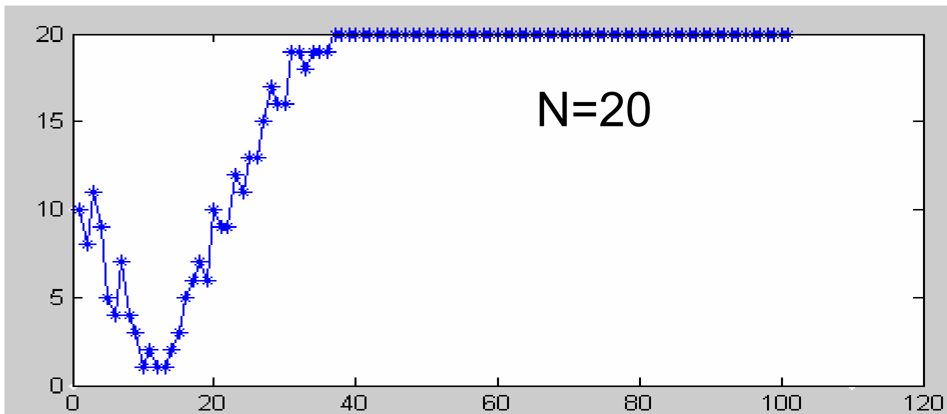
Simulation

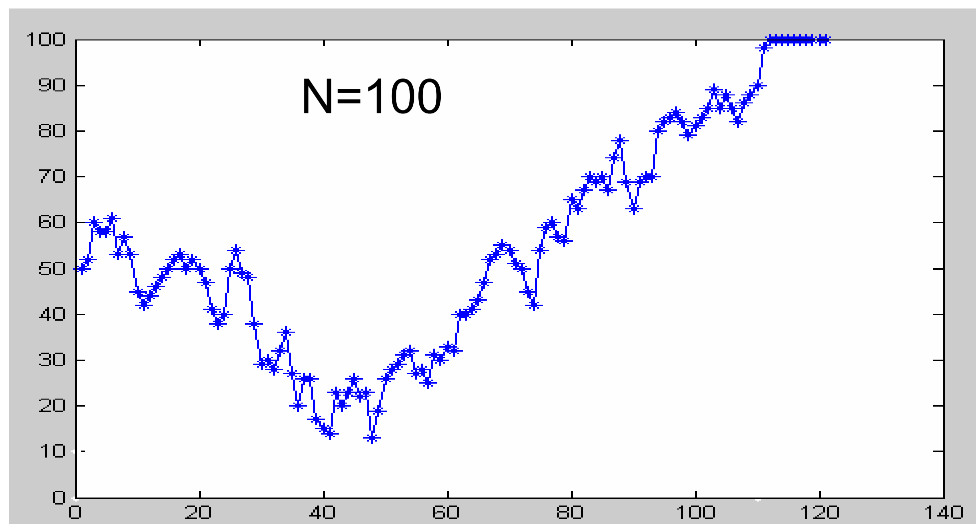
Some sample runs with 10 particles



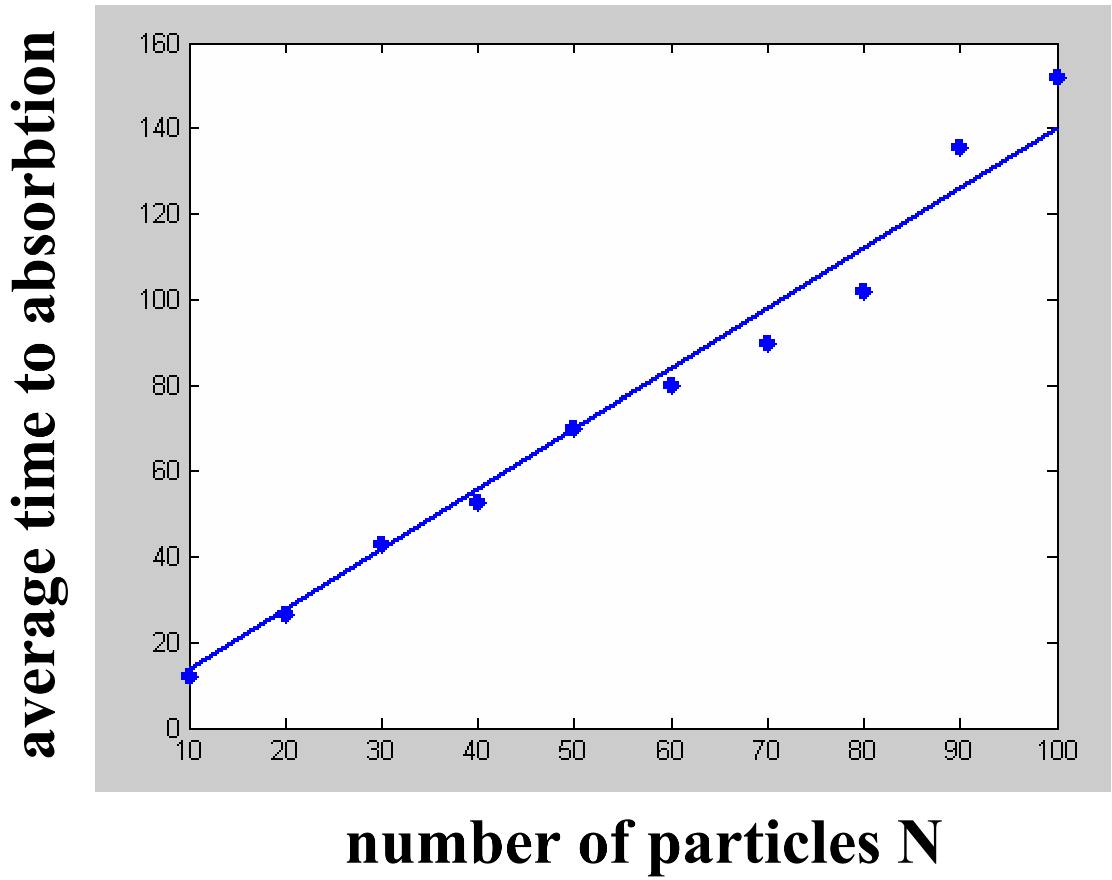
More Sample Runs







Average Time to Absorbtion



Dots - from running simulator (100 trials at N=10,20,30...)
Line - plot of 1.4 N, the asymptotic analytic estimate (King and Forsyth)

More Generally

Implications of stationary process model with no noise, in a discrete state space.

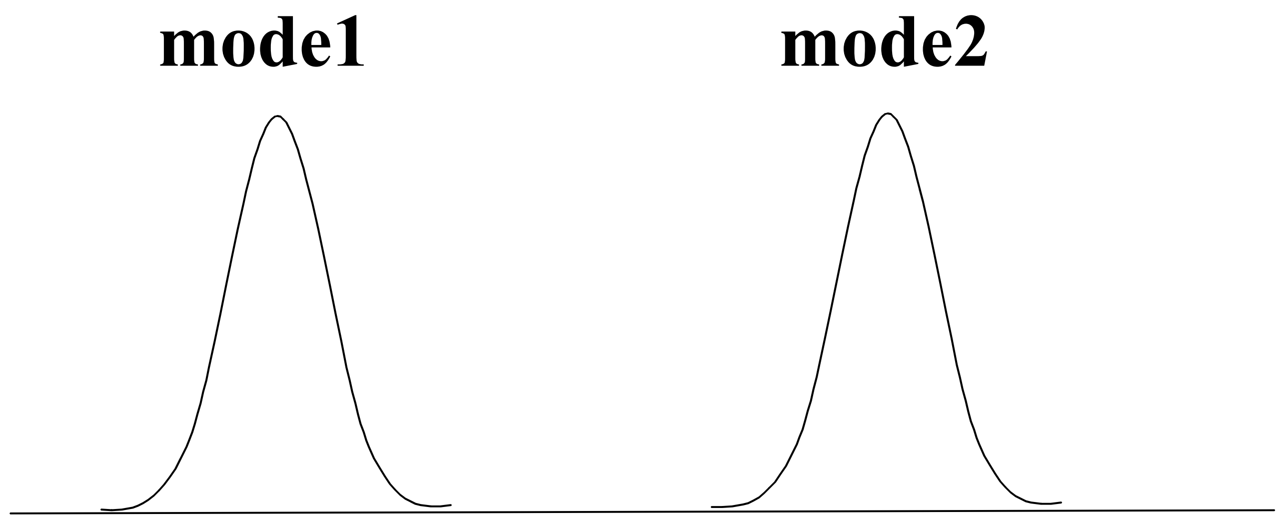
- any time any bucket contains zero particles, it will forever after have zero particles (for that run).

- there is typically a nonzero probability of getting zero particles in a bucket sometime during the run.

- thus, over time, the particles will inevitably collapse into a single bucket.

Extending to Continuous Case

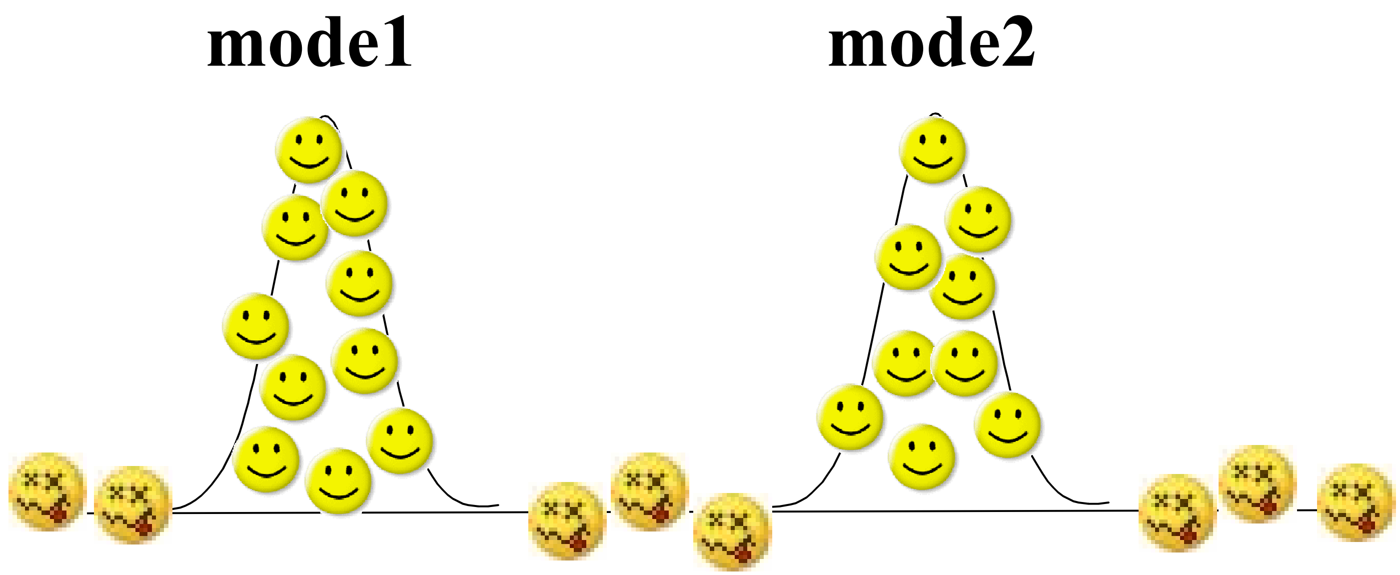
A similar thing happens in more realistic cases. Consider a continuous case with two stationary modes in the likelihood, and where each mode has small variance with respect to distance between modes.



Robert Collins
CSE598C

Extending to Continuous Case

The very low variance between modes is fatal to any particles that try to cross from one to the other via diffusion.



Extending to Continuous Case

Each mode thus becomes an isolated island, and we can reduce this case to our previous two-state analysis (each mode is one discrete state)