Economics 403A

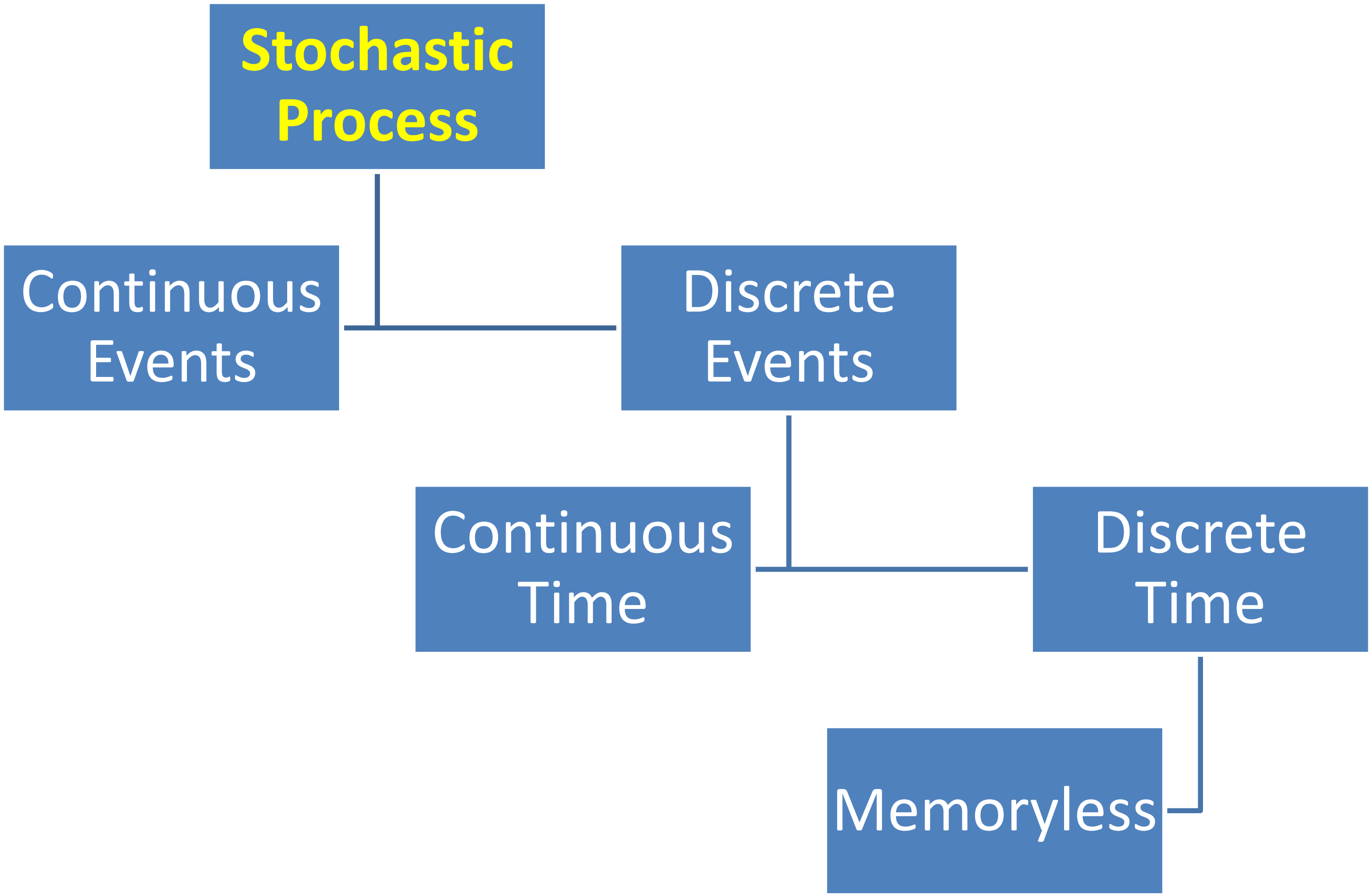
Markov Chains

Dr. Randall R. Rojas

Today's Class

- Historical Background
- Introduction to Markov Chains
  - Two-State Example
  - Three-State Example
- Properties of Markov Chains
- Types of Markov Chains
- Theorems
- MCMC
- Gibbs Sampling Algorithm

Stochastic Processes



Historical Background

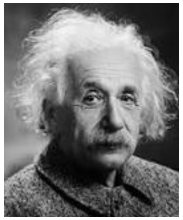
- **Andrey Markov** (Chebyshev was his advisor)

  - 1906 –First paper on the subject of Markov Chains

- Paul Eherenfest: Diffusion model, Early 1900s

  - Statistical interpretation of the second law of thermodynamics: The entropy of a closed system can only increase. Proposed the “Urn Model” to explain diffusion.

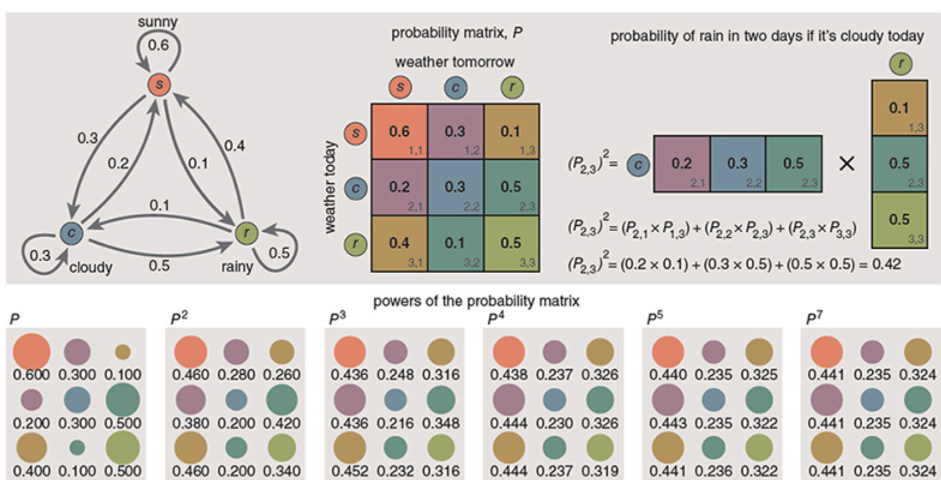
- Albert Einstein: Brownian Motion, 1905



  - Realized Brownian (after Robert Brown 1827) motion would provide a magnifying glass into the world of the atom. Brownian motion has been extensively modeled by Markov Chains.

Introduction

- A **Markov Chain** is a mathematical model that describes a set of states and transitions between them.
  - Each transition probability is the probability of moving from one state to another in one step.
  - The transition probabilities are independent of the past, and depend only on the two states involved. The matrix of transition probabilities is called the transition matrix.



Introduction

- A **Markov Chain** consists of three elements:
  - A State Space $\boldsymbol{x}$ = set of values the chain can take
  - Note: Your textbook uses **S** = State Space

  - A Transition Probability $\boldsymbol{p(x^{t+1}|x^{t})}$ =probability of moving from state x^t to x^{t+1}
  - In general, $\boldsymbol{p_{ij}}$ = conditional probability of being in state x_i at step $t+1$ given that the process was in state x_j at step t.

  - Initial condition distribution $\boldsymbol{\pi(0)}$= probability of being in any one of the possible states at $t=0$.

Introduction

- Weather Forecast Example:
  - Assume on any given day it is either *sunny* (S)or *rainy* (R). Therefore, our state space $\boldsymbol{x} = \{\boldsymbol{S}, \boldsymbol{R}\}$
  - Historical meteorological data reveal that:

$$
\left.
\begin{aligned}
p(x^{Tomorrow} = S \mid x^{Today} = S) = 0.7 \\
p(x^{Tomorrow} = S \mid x^{Today} = R) = 0.2 \\
p(x^{Tomorrow} = R \mid x^{Today} = S) = 0.3 \\
p(x^{Tomorrow} = R \mid x^{Today} = R) = 0.8
\end{aligned}
\right\}
\boldsymbol{P} = \begin{array}{cc} & \begin{array}{cc} S^{t+1} & R^{t+1} \end{array} \\ \begin{array}{c} S^t \\ R^t \end{array} & \underbrace{\begin{pmatrix} 0.7 & 0.3 \\ 0.2 & 0.8 \end{pmatrix}}_{\text{Transition Probability}} \end{array}
$$

  - Initial State: Today is a sunny day→ $\boldsymbol{x^0} = [1\ 0]$
  - Note: 1 = 100% sunny and 0% rainy

Introduction

- Q1: What is the probability that it will be sunny tomorrow given that today is sunny?

$\rightarrow x^1 = x^0 P = [1\ 0] \begin{pmatrix} 0.7 & 0.3 \\ 0.2 & 0.8 \end{pmatrix} = [0.7\ 0.3]$

Conclusion: There is a 70% that it will be sunny tomorrow given that it sunny today, and a 30% that it will be rainy.

Introduction

- Q2: What is the probability that it will be sunny two days from now given that today is sunny?

$\rightarrow x^2 = x^1 P = (x^0 P)P = x^0 P^2$

$$= [1\ 0] \begin{pmatrix} 0.7 & 0.3 \\ 0.2 & 0.8 \end{pmatrix} \begin{pmatrix} 0.7 & 0.3 \\ 0.2 & 0.8 \end{pmatrix}$$

$$= [0.55\ 0.45]$$

Conclusion: There is a 55% that it will be sunny two days from now given that it sunny today, and a 45% that it will be rainy.

Introduction

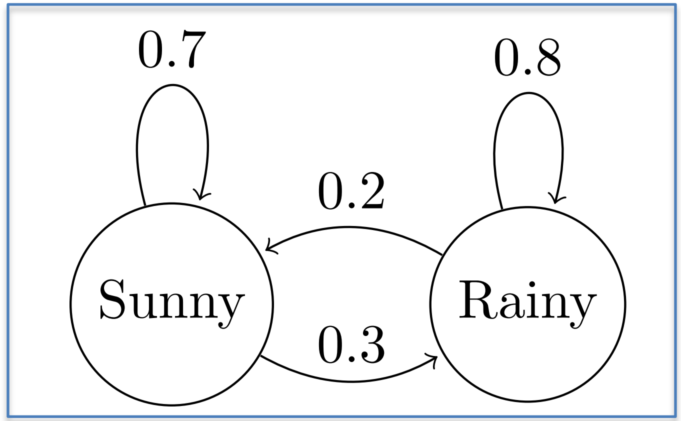
- Q3: What is the probability that it will be sunny three weeks from now given that today is sunny?

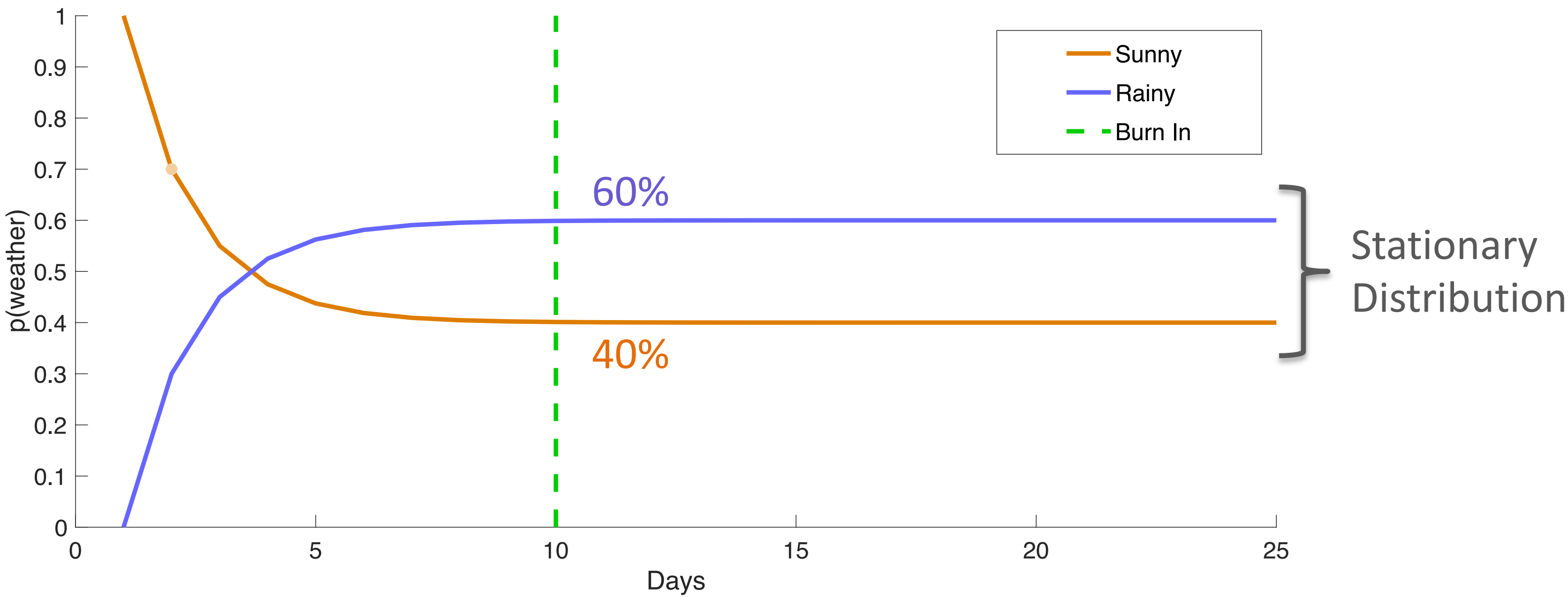
$$\rightarrow x^{21} = x^{0} P^{21} = [1\ 0] \begin{pmatrix} 0.4 & 0.6 \\ 0.4 & 0.6 \end{pmatrix}$$

$$= [0.40\ 0.60]$$

Conclusion: There is a 40% that it will be sunny two days from now given that it sunny today, and a 60% that it will be rainy.

Markov Model Representation





The **burn in** period for the Markov chain is the number of transitions it takes the chain to move from the initial conditions to the stationary distribution.

Introduction

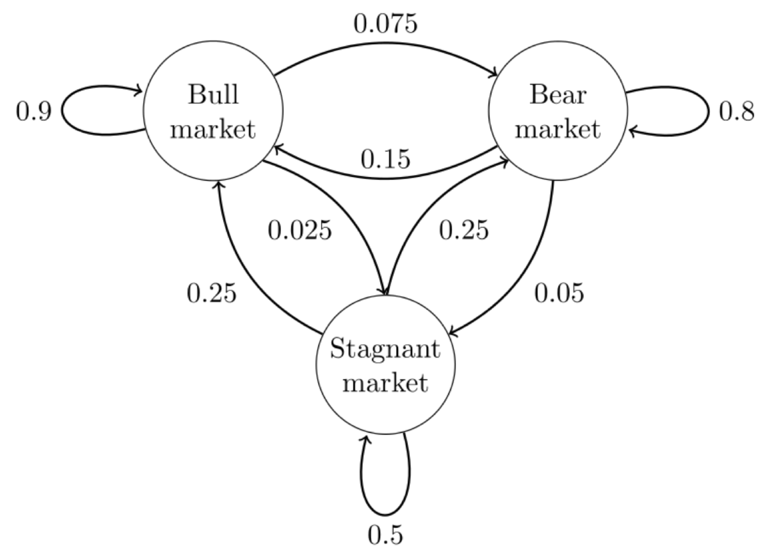
- Financial Market Example:
  - Assume on any given day the market is either bullish, bearish, or stagnant. Therefore, our state space

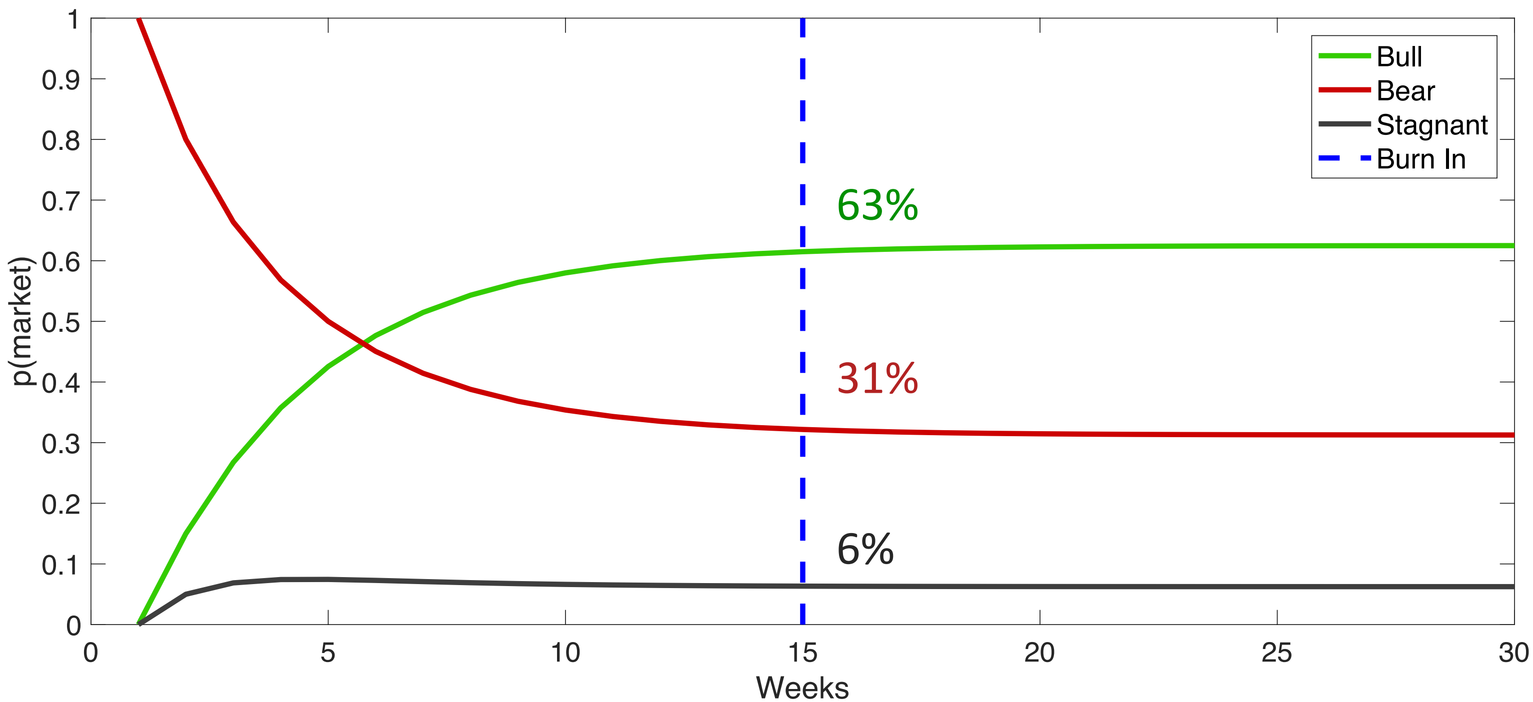
$x = \{Bull, Bear, Stagnant\}$

P (To) / (From)	**Bull**	**Bear**	**Stagnant**
Bull	0.900	0.075	0.025
Bear	0.150	0.800	0.050
Stagnant	0.250	0.250	0.500

Initial State: Today the market is Bearish → $x^0 = [0\ 1\ 0]$
→Find the steady state solution (Burn in Time)

Markov Model Representation





Introduction

- A sequence of trials of an experiment is a Markov Chain if:
  1. The outcome of each experiment is one of a set of discrete states;
  2. The outcome of an experiment depends only on the present state, and not on any past states.

- If the time parameter is discrete $\{t_1, t_2, t_3, \ldots\ldots\}$, it is called Discrete Time Markov Chain (DTMC).

- If time parameter is continues, ($t \geq 0$) it is called Continuous Time Markov Chain (CTMC).

Properties of Markov Chains

Transition Matrix

$$P = pij = \begin{array}{cc} & \overbrace{\begin{array}{ccc} S_1 & S_2 & S_3 \end{array}}^{\text{State } t} \\ \text{State } t+1 \left\{ \begin{array}{c} S_1 \\ S_2 \\ S_3 \end{array} \right. & \begin{pmatrix} p_{11} & p_{12} & p_{13} \\ p_{21} & p_{22} & p_{23} \\ p_{31} & p_{32} & p_{33} \end{pmatrix} \end{array}$$

- It is square matrix , since all possible states must be used both as rows and as columns.
- All entries are between 0 and 1, because all entries represent probabilities.
- The sum of the entries in any row must be 1, since the numbers in the row give you the probability of changing from the state at the left to one of the states indicated across the top.

Types of Markov Chains

- Def: Regular Markov Chains
  - A Markov chain is a **Regular Markov** chain if some power of the transition matrix has only positive entries. That is, if we define the (i, j) entry of P^n to be $p^n{}_{ij}$, then the Markov chain is regular if there is some n such that $p^n{}_{ij} > 0$ for all (i,j).

Example (1)

$$P = \begin{pmatrix} 0.4 & 0.6 \\ 0.4 & 0.6 \end{pmatrix}$$

All entries are positive

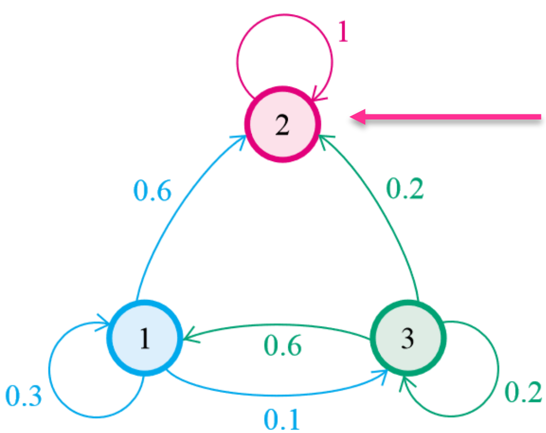
Example (2)

$$P = \begin{pmatrix} 1/2 & 1/4 & 1/4 \\ 1/2 & 0 & 1/2 \\ 1/4 & 1/4 & 1/2 \end{pmatrix}$$

If you compute P^2, all entries are positive

Types of Markov Chains

- Def: Absorbing Markov Chains
  - A state S_k of a Markov chain is called an **Absorbing State** if, once the Markov chains enters the state, it remains there forever.
  - A Markov chain is called an absorbing chain if It has at least one absorbing state.
  - In an absorbing Markov chain, a state which is not absorbing is called **Transient**.
  - For every state in the chain, the probability of reaching an absorbing state in a finite number of steps is nonzero.



State 2 is an absorbing state

$$P = \begin{pmatrix} 0.3 & 0.6 & 0.1 \\ 0 & \mathbf{1} & 0 \\ 0.6 & 0.2 & 0.2 \end{pmatrix}$$

$P_{22} = 1$, $P^2_{22}=1$, …,$P^n_{22}=1$

Types of Markov Chains

- Def: Irreducible (or Ergodic) Markov Chains
  - A Markov chain is called an **irreducible** or **ergodic** chain if it is possible to go from every state to every state (not necessarily in one move).
  - i and j communicate if they are accessible from each other. This is written $i \leftrightarrow j$

Example (1)

$$P = \begin{array}{c c} & \begin{array}{cc} S_1 & S_2 \end{array} \\ \begin{array}{c} S_1 \\ S_2 \end{array} & \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \end{array}$$

This chain is ergodic since it is possible to move from any state to any other state

Example (2) Ehrenfest's urn

$$\mathbf{P} = \begin{array}{c c} & \begin{array}{ccccc} 0 & 1 & 2 & 3 & 4 \end{array} \\ \begin{array}{c} 0 \\ 1 \\ 2 \\ 3 \\ 4 \end{array} & \begin{pmatrix} 0 & 1 & 0 & 0 & 0 \\ 1/4 & 0 & 3/4 & 0 & 0 \\ 0 & 1/2 & 0 & 1/2 & 0 \\ 0 & 0 & 3/4 & 0 & 1/4 \\ 0 & 0 & 0 & 1 & 0 \end{pmatrix} \end{array}$$

In this example, if we start in state 0 we will, after any even number of steps, be in either state 0, 2 or 4, and after any odd number of steps, be in states 1 or 3. Thus this chain is ergodic but not regular.

Classification of States (Summary)

Let p_{jj} be the probability of returning to state j after leaving j.

- A state j is said transient if $p_{jj} < 1$
- A state j is said recurrent if $p_{jj} = 1$
- A state j is said absorbing if $p_{jj} = 1$
- A state j is said periodic if it is only visited in a number of steps which is multiple of an integer $m > 1$, called the period. A state j is said aperiodic otherwise
- A Markov chain is aperiodic if the period of each state is equal to 1
- A state j is said reachable from a state i if there is a path from i to j in the state transition diagram.
- A subset S of states is said closed if there is no transition leaving S.

Let T_{jj} be the average recurrent time, i.e., the time it takes to return to j

- A recurrent state j is positive recurrent if $E[T_{jj}]$ is finite.
- A recurrent state j is null recurrent if $E[T_{jj}] = \infty$.

Theorems

1. If a Markov chain has a finite state space, then at least one state is recurrent.
2. If i is a recurrent state and j is reachable from i, then state j is recurrent.
3. If S is a finite closed irreducible set of states, then every state in S is recurrent.
4. If i is a positive recurrent state and j is reachable from i, then state j is positive recurrent.
5. If S is a closed irreducible set of states, then every state in S is positive recurrent or every state in S is null recurrent or every state in S is transient.
6. If S is a finite closed irreducible set of states, then every state in S is positive recurrent.

Markov Chain Monte Carlo (MCMC)

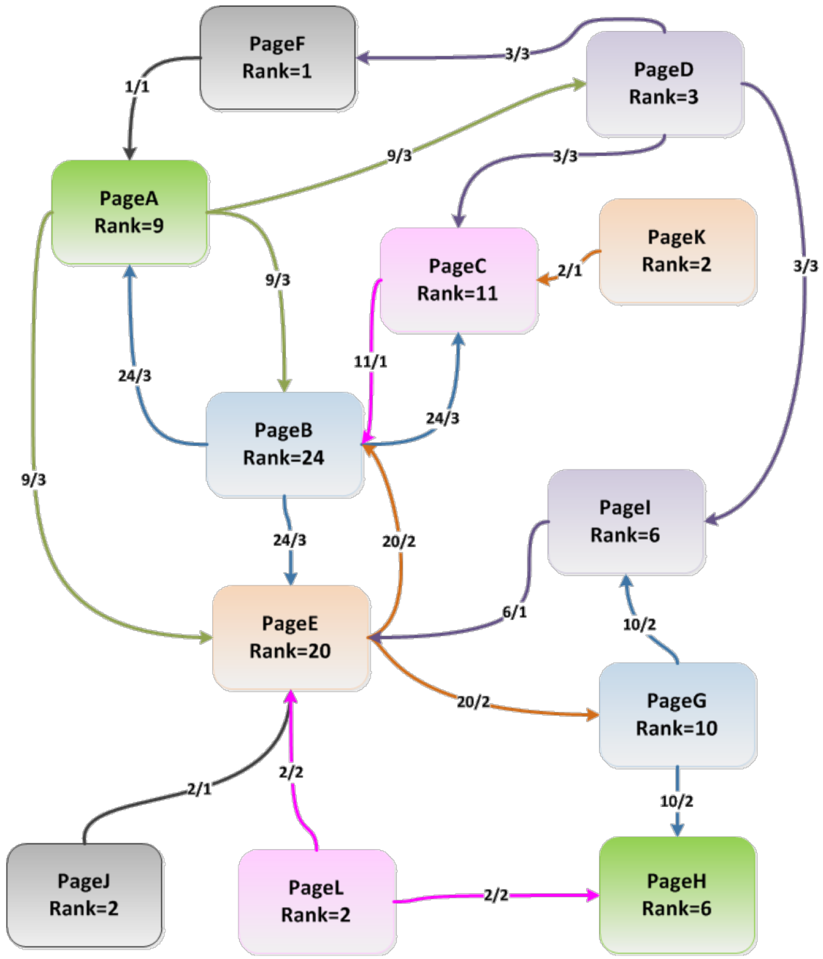
- Design a Markov Chain on finite state space

$$\text{state space}: x^{(i)} \in \{x_1, x_2, \ldots, x_s\}$$

$$\text{Markov property}: \; p(x^{(i)} \mid x^{(i-1)}, \ldots, x^{(1)}) = T(x^{(i)} \mid x^{(i-1)})$$

…such that when simulating a trajectory of states from it, it will explore the state space spending more time in the most important regions (i.e. where p(x) is large)

Markov Chain Monte Carlo (MCMC)



- Supposing you browse this for infinitely long time, what is the probability to be at page x_i.
- No matter where you started off.

Pagerank Algorithm

Markov Chain Monte Carlo (MCMC)

Google vs. MCMC

- Google is given **T** and finds $p(x)$
- MCMC is given $p(x)$ and finds **T**
- But it also needs a transition probability distribution to be specified.
- Q: Do all MCs have a stationary distribution? →No

Markov Chain Monte Carlo (MCMC)

Conditions for existence of a unique stationary distribution

- MCMC samplers are irreducible and aperiodic MCs that converge to the target distribution
- These 2 conditions are not easy to impose directly
- Reversibility (also called 'detailed balance') is a sufficient (but not necessary) condition for $p(x)$ to be the stationary distribution.

$$p(x^t)T(x^{t-1}|x^t) = p(x^{t-1})T(x^t|x^{t-1})$$

$$\rightarrow p(x^t) = \sum_{x^{t-i}} p(x^{t-1})T(x^t|x^{t-1})$$

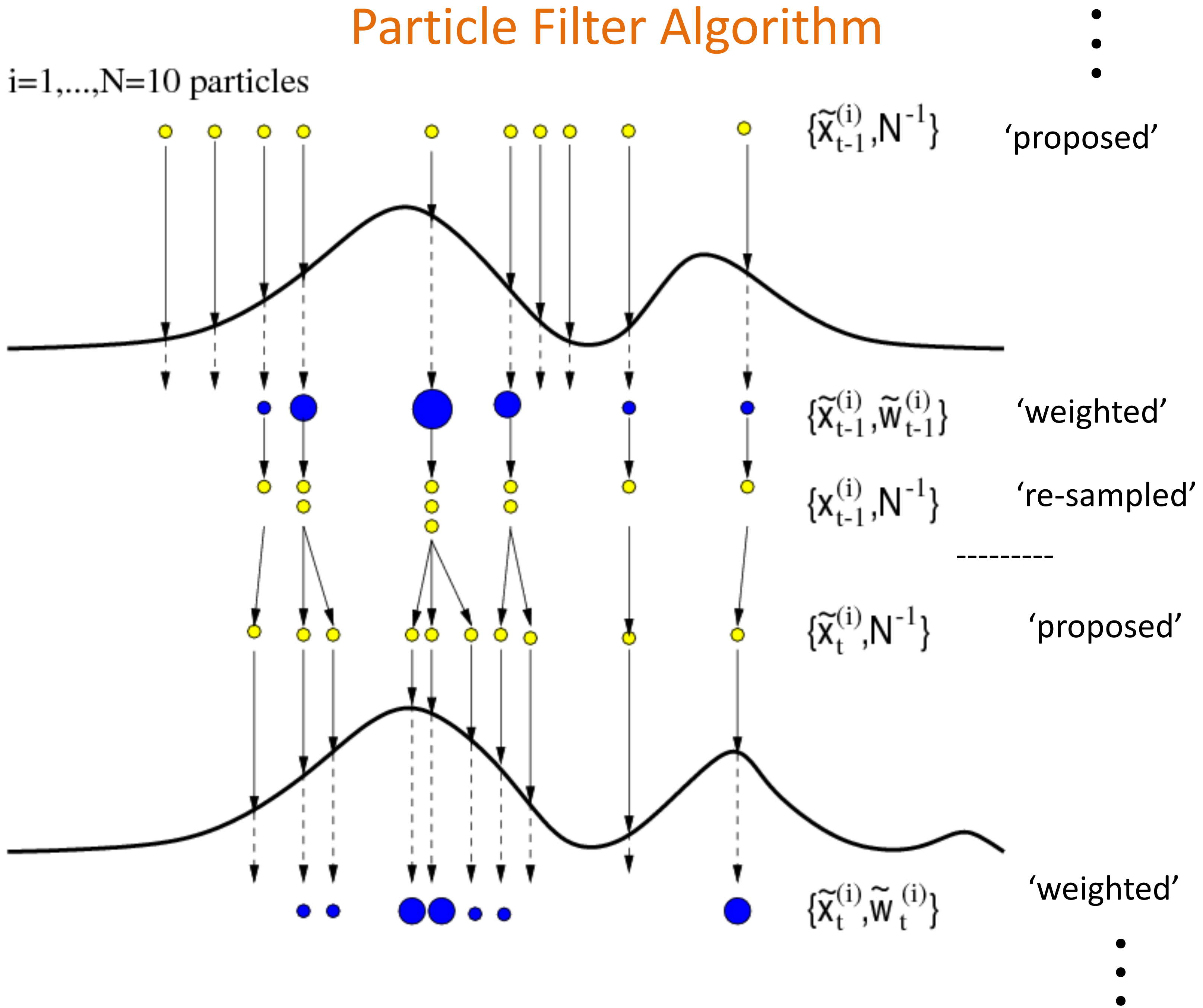
  - It is easier to work with this condition

Markov Chain Monte Carlo (MCMC)

MCMC Algorithms

- Metropolis-Hastings
- Metropolis
  - Mixtures and Blocks
- Gibbs Sampling
- Sequential Monte Carlo & Particle Filters
- Others
- Auxiliary variable samplers
  - Hybrid Monte Carlo
  - Tries to avoid ‘random walk’ behavior, i.e. to speed up convergence
- Reversible jump MCMC
  - For comparing models of different dimensionalities (in ‘model selection’ problems)
- Adaptive MCMC

Particle Filter Algorithm



Gibbs Sampling Algorithm

- Suppose $p(x,y)$ is a p.d.f. or p.m.f. that is difficult to sample from directly.
- Suppose, though, that we can easily sample from the conditional distributions $p(x|y)$ and $p(y|x)$.
- The Gibbs sampler proceeds as follows:
  1. Set x and y to some initial starting values
  2. Sample $x|y$, then sample $y|x$, then $x|y$, and so on.

Gibbs Sampling Algorithm

1. Set (x_0, y_0) to some starting value.
2. Sample $x_1 \sim p(x|y_0)$, i.e., from the conditional distribution $X|Y=y_0$

   Current state: (x_1, y_0)

   Sample $y_1 \sim p(y|x_1)$, i.e., from the conditional distribution $Y|X=x_1$

   Current state: (x_1, y_1)
3. Sample $x_2 \sim p(x|y_1)$, i.e., from the conditional distribution $X|Y=y_1$

   Current state: (x_2, y_1)

   Sample $y_2 \sim p(y|x_2)$, i.e., from the conditional distribution $Y|X = x_2$

   Current state: (x_2, y_2)

   ⋮

Repeat iterations 2 and 3, N times

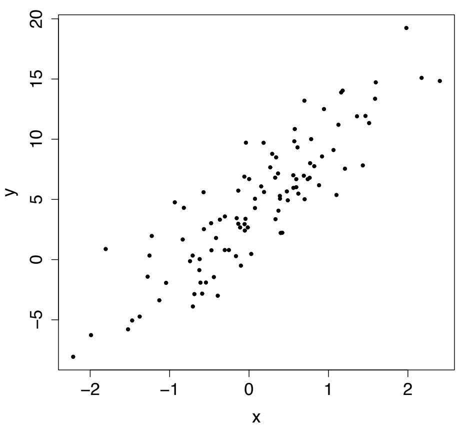
Gibbs Sampling Algorithm

- Once completed, This procedure defines a sequence of pairs of random variables $(X_0, Y_0), (X_1, Y_1), (X_2, Y_2), (X_3, Y_3), \ldots$ that satisfy the property of being a Markov chain.
- The conditional distribution of (X_i, Y_i) given all of the previous pairs depends only on (X_{i-1}, Y_{i-1})
- (x_0,y_0) chosen to be in a region of high probability under $p(x,y)$, but often this is not so easy.
- We run the chain for N iterations and discard the first B samples $(X_1,Y_1),\ldots,(X_B,Y_B)$. This is called burn-in.
  - Typically: if you run the chain long enough, the choice of B does not matter.
- The performance of an MCMC algorithm, i.e., how quickly the sample averages converge, is referred to as the mixing rate.

Gibbs Sampling Algorithm

- Regression Example
  - R Code: https://github.com/stablemarkets/BayesianTutorials
  - Article: https://www.r-bloggers.com/bayesian-simple-linear-regression-with-gibbs-sampling-in-r/

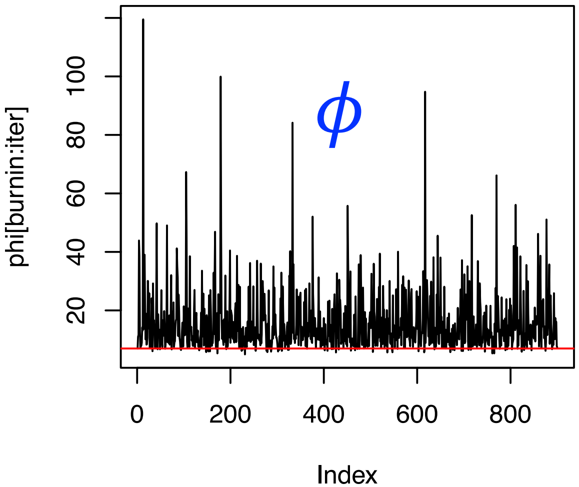
- Model: $y_i \sim N(\beta_0 + \beta_1 x_i, \phi)$
- Estimate β_0 and β_1
- Assume Normal priors on β_0 and β_1
- Assume Inverse Gamma on the variance ϕ

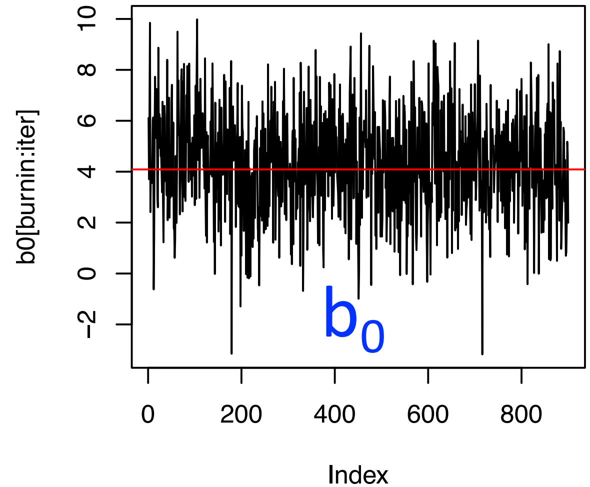


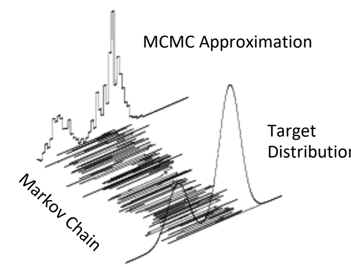
$$
\begin{gathered}
y_i|\beta_0, \beta_1, \phi \sim N(\beta_0 + \beta_1 x_i, \phi) \\
\beta_0|\mu_0, \tau_0 \sim N(\mu_0, \tau_0) \\
\beta_1|\mu_1, \tau_1 \sim N(\mu_1, \tau_1) \\
\phi|\alpha, \gamma \sim IG(\alpha, \gamma)
\end{gathered}
$$

Bayesian Model

We will use Gibbs sampling to produces samples from the posterior distribution of each parameter of interest







Note: LS estimates are b_0 = 4.1 and b_1= 5.2