Bayesian Techniques for Parameter Estimation

"He has Van Gogh's ear for music," Billy Wilder

Reading: Sections 4.6, 4.8 and Chapter 12

Statistical Inference

Goal: The goal in statistical inference is to make conclusions about a phenomenon based on observed data.

Frequentist: Observations made in the past are analyzed with a specified model. Result is regarded as confidence about state of real world.

- Probabilities defined as frequencies with which an event occurs if experiment is repeated several times.
- Parameter Estimation:
    - o Relies on estimators derived from different data sets and a specific sampling distribution.
    - o Parameters may be unknown but are fixed and deterministic.

Bayesian: Interpretation of probability is subjective and can be updated with new data.

- Parameter Estimation: Parameters are considered to be random variables having associated densities.

Bayesian Inference

Framework:

- Prior Distribution: Quantifies prior knowledge of parameter values.
- Likelihood: Probability of observing a data if we have a certain set of parameter values; Comes from observation models in Chapter 5!
- Posterior Distribution: Conditional probability distribution of unknown parameters given observed data.

Joint PDF: Quantifies all combination of data and observations

$$\pi(\theta, y) = \pi(y|\theta)\pi_0(\theta)$$

Bayes' Relation: Specifies posterior in terms of likelihood, prior, and normalization constant

$$\pi(\theta|y) = \frac{f(y|\theta)\pi_0(\theta)}{\int_{\mathbb{R}^p} f(y|\theta)\pi_0(\theta)d\theta}$$

Problem: Evaluation of normalization constant typically requires high dimensional integration.

Bayesian Inference

Uninformative Prior: No *a priori* information parameters

$$\text{e.g., } \pi_0(\theta) = 1$$

Informative Prior: Use conjugate priors; prior and posterior from same distribution

$$\pi(\theta|y) = \frac{f(y|\theta)\pi_0(\theta)}{\int_{\mathbb{R}^p} f(y|\theta)\pi_0(\theta)d\theta}$$

Evaluation Strategies:

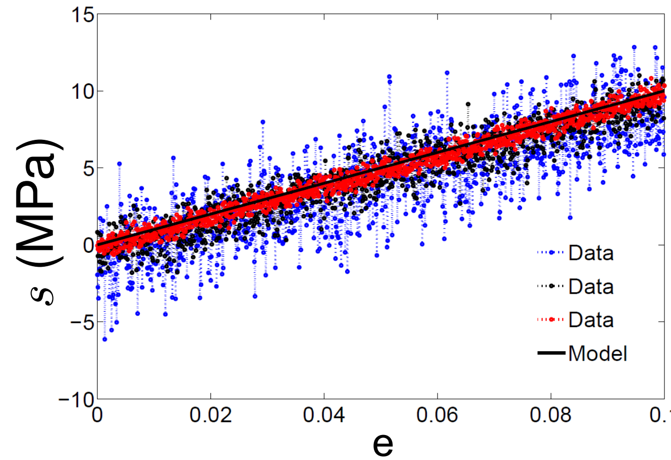
- Analytic integration --- Rare
- Classical Gaussian quadrature; e.g., p = 1 - 4
- Sparse grid quadrature techniques; e.g., p = 5 - 40
- Monte Carlo quadrature Techniques
- Markov chain methods

Bayesian Inference: Motivation

Example: Displacement-force relation (Hooke's Law)

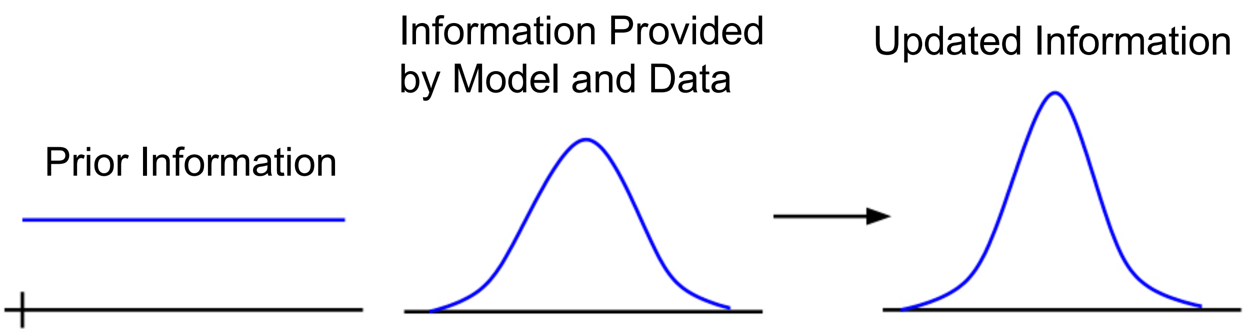
$$s_i = Ee_i + \varepsilon_i \ , \ i = 1, \ldots, N$$

$$\varepsilon_i \sim N(0, \sigma^2)$$



Parameter: Stiffness E

Strategy: Use model fit to data to update prior information



$\pi_0(E)$ $\quad e^{-\sum_{i=1}^{N}[s_i - Ee_i]^2/2\sigma^2}$ $\quad \pi(E|s)$

Data Model

Non-normalized Bayes' Relation:

$$\pi(E|s) = e^{-\sum_{i=1}^{N}[s_i - Ee_i]^2/2\sigma^2}\pi_0(E)$$

Bayesian Inference

Bayes' Relation: Specifies posterior in terms of likelihood and prior

Likelihood: $e^{-\sum_{i=1}^{N}[s_i - Ee_i]^2/2\sigma^2}$, $q = E$

$\upsilon = [s_1, \ldots, s_N]$

Posterior Distribution

$$\pi(\theta|y) = \frac{f(y|\theta)\pi_0(\theta)}{\int_{\mathbb{R}^p} f(y|\theta)\pi_0(\theta)d\theta}$$

Prior Distribution

Normalization Constant

- Prior Distribution: Quantifies prior knowledge of parameter values
- Likelihood: Probability of observing a data given set of parameter values.
- Posterior Distribution: Conditional distribution of parameters given observed data.

Problem: Can require high-dimensional integration

- e.g., Many applications: p = 10-50!
- Solution: Sampling-based Markov Chain Monte Carlo (MCMC) algorithms.
- Metropolis algorithms first used by nuclear physicists during Manhattan Project in 1940's to understand particle movement underlying first atomic bomb.

Bayesian Model Calibration

Bayes' Relation:

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$

Bayesian Model Calibration:

- Parameters assumed to be random variables

$$\pi(\theta|y) = \frac{f(y|\theta)\pi_0(\theta)}{\int_{\mathbb{R}^p} f(y|\theta)\pi_0(\theta)d\theta}$$

Example: Coin Flip

$$Y_i(\omega) = \begin{cases} 0 & , \quad \omega = T \\ 1 & , \quad \omega = H \end{cases}$$

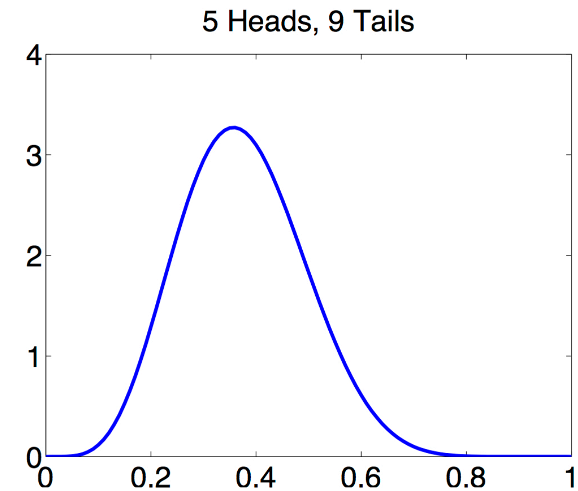
Likelihood:

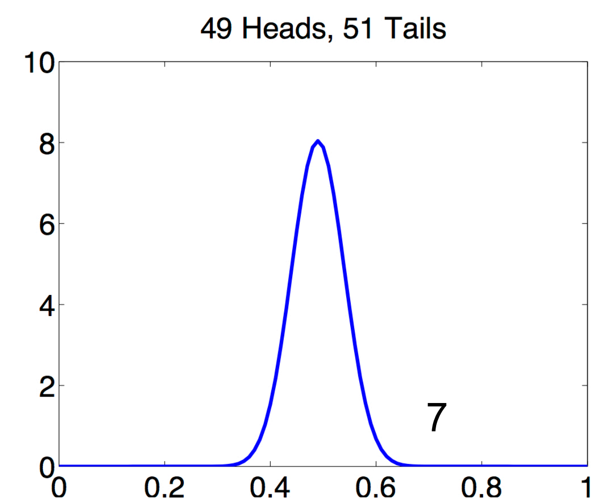
$$\pi(y|\theta) = \prod_{i=1}^{N} \theta^{y_i}(1-\theta)^{1-y_i}$$

$$= \theta^{N_1}(1-\theta)^{N_0}$$

Posterior with flat Prior: $\pi_0(\theta) = 1$

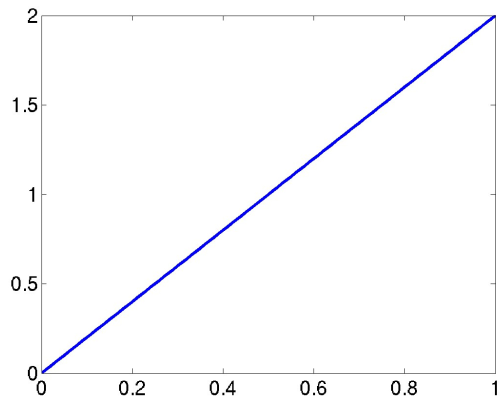
$$\pi(\theta|y) = \frac{\theta^{N_1}(1-\theta)^{N_0}}{\int_0^1 \theta^{N_1}(1-\theta)^{N_0} dq} = \frac{(N+1)!}{N_0!N_1!}\theta^{N_1}(1-\theta)^{N_0}$$



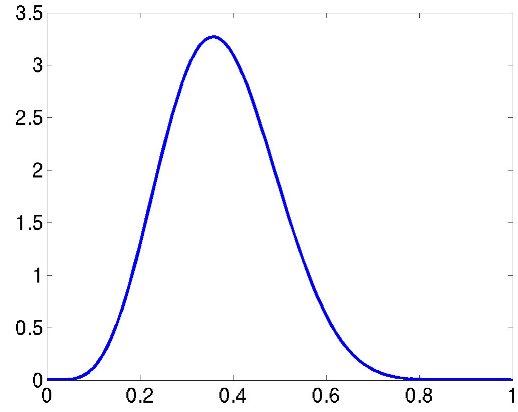


Bayesian Inference

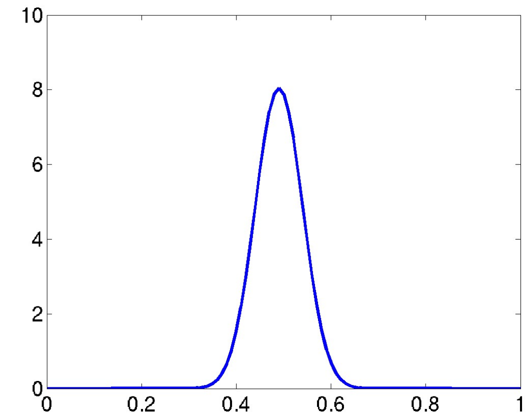
Example:



1 Head, 0 Tails



5 Heads, 9 Tails



49 Heads, 51 Tails

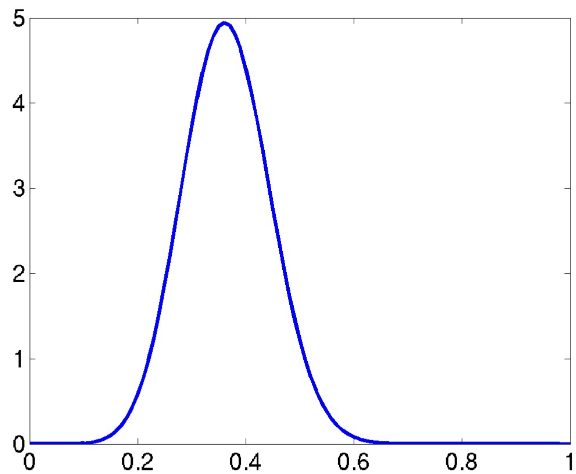
Note: For $N = 1$, frequentist theory would give probability 1 or 0

Bayesian Inference

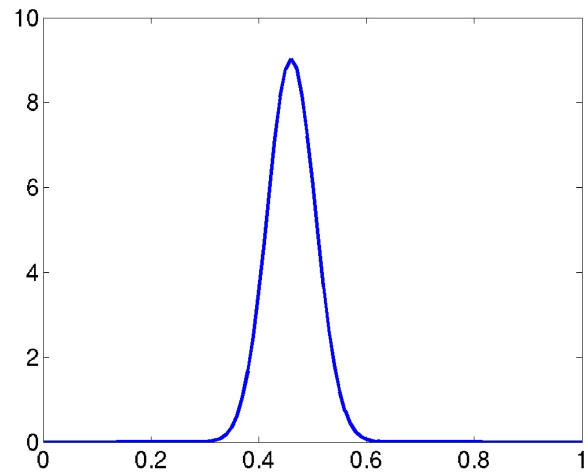
Example: Now consider

$$\pi_0(\theta) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(\theta-\mu)^2/2\sigma^2}$$

with $\mu = 0.3$ and $\sigma = 0.1$



5 Heads, 5 Tails



50 Heads, 50 Tails

Note: Poor informative prior incorrectly influences results for a long time.

Parameter Estimation Problem

Observation Model:

$$y_i = f_i(\theta) + \varepsilon_i \ , \ \ i = 1, \ldots, n$$

Assumption: Assume that measurement errors are iid and $\varepsilon_i \sim N(0, \sigma^2)$

Likelihood:

$$f(y|\theta) = L(\theta, \sigma|y) = \frac{1}{(2\pi\sigma^2)^{n/2}} e^{-SS_\theta / 2\sigma^2}$$

where

$$SS_\theta = \sum_{j=1}^{n} [y_j - f_i(\theta)]^2$$

is the sum of squares error.

Parameter Estimation: Example

Example: Consider the spring model

$$\ddot{z} + C\dot{z} + Kz = 0$$

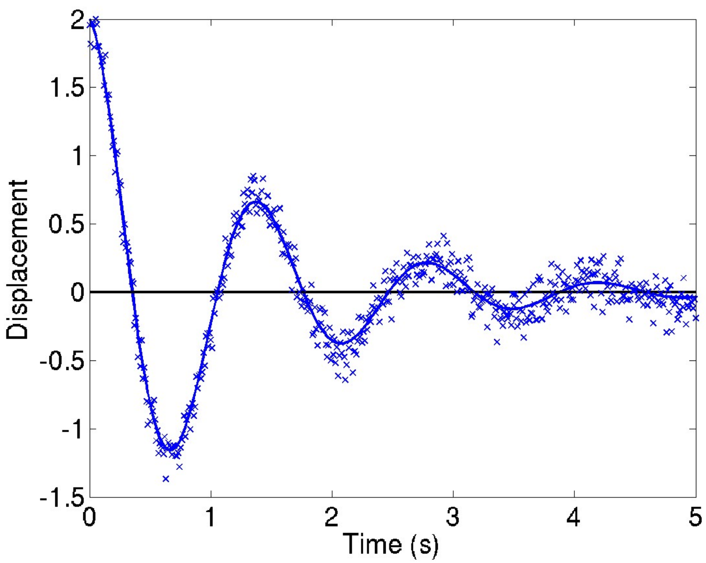
$$z(0) = 2 \ , \ \dot{z}(0) = -C$$

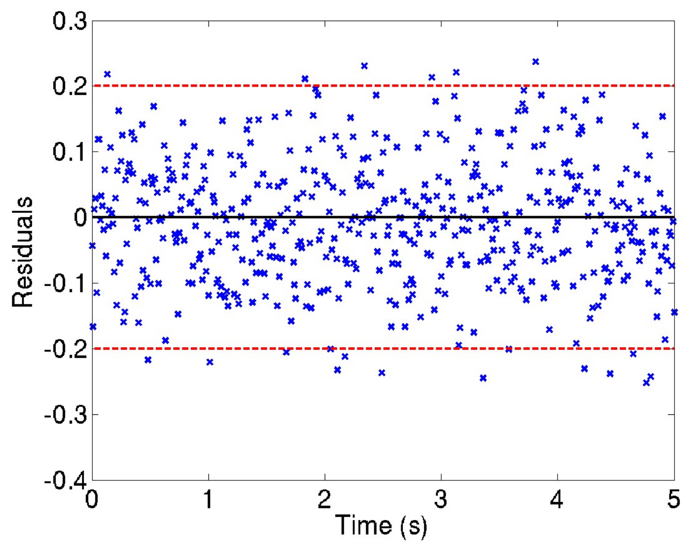
Note: Take $K = 20.5$, $C^0 = 1.5$

which has the solution

$$z(t) = 2e^{-Ct/2} \cos(\sqrt{K - C^2/4} \cdot t)$$

Take K to be known and $\theta = C$. Assume that $\varepsilon_i \sim N(0, \sigma_0^2)$ where $\sigma_0 = 0.1$





Parameter Estimation: Example

Example: The sensitivity matrix is

$$\mathcal{X}(\theta) = \left[\frac{\partial y}{\partial C}(t_1,\theta),\cdots,\frac{\partial y}{\partial C}(t_n,\theta)\right]^T$$

where

$$\frac{\partial y}{\partial C} = e^{-Ct/2}\left[\frac{Ct}{\sqrt{4K-C^2}}\sin\left(\sqrt{K-C^2/4}\cdot t\right) - t\cos\left(\sqrt{K-C^2/4}\cdot t\right)\right]$$

Here

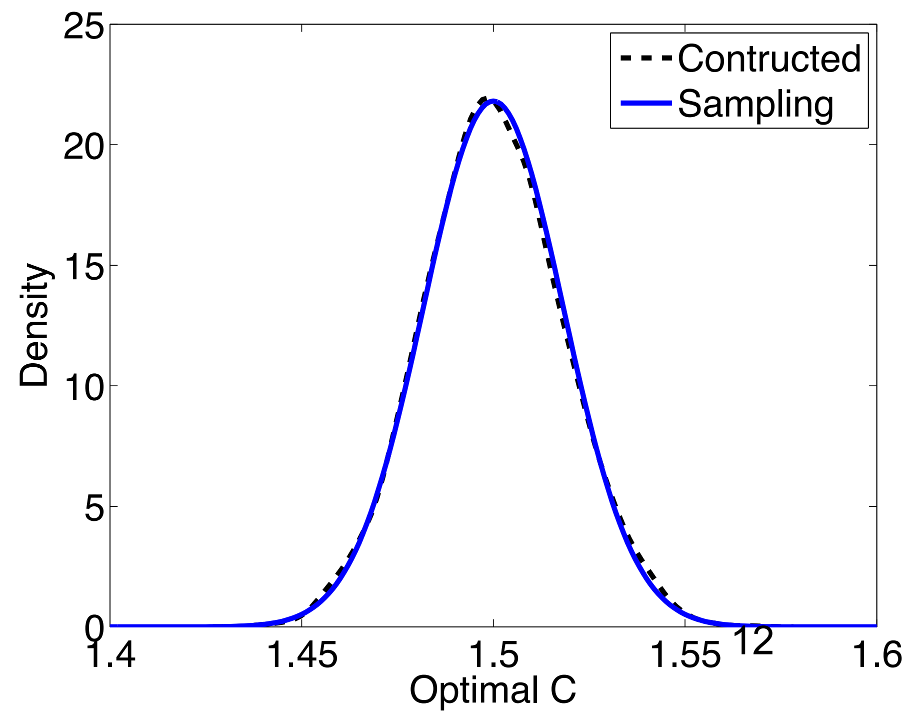
$$V = \sigma_c^2 = \sigma_0^2\left[\mathcal{X}^T(\theta)\mathcal{X}(\theta)\right]^{-1} = 3.35\times 10^{-4}$$

so that

$$\widehat{C} \sim N\left(C_0,\sigma_c^2\right)\ ,\ \sigma_c = 0.0183$$

Note: In 10,000 simulations, 9455 of confidence intervals contained true parameter value.

Figure: Sampling distribution compared with that constructed using 10,000 estimated values of C.



Parameter Estimation: Example

Bayesian Inference: Employ the flat prior

$$\pi_0(\theta) = \chi_{[0,\infty)}(\theta)$$

Posterior Distribution:

$$\pi(\theta|y) = \frac{e^{-SS_\theta/2\sigma_0^2}}{\int_0^\infty e^{-SS_\zeta/2\sigma_0^2} d\zeta} = \frac{1}{\int_0^\infty e^{-(SS_\zeta - SS_\theta)/2\sigma_0^2} d\zeta}$$

Issue: $e^{-SS_{\theta_{MAP}}} \approx 3 \times 10^{-113}$

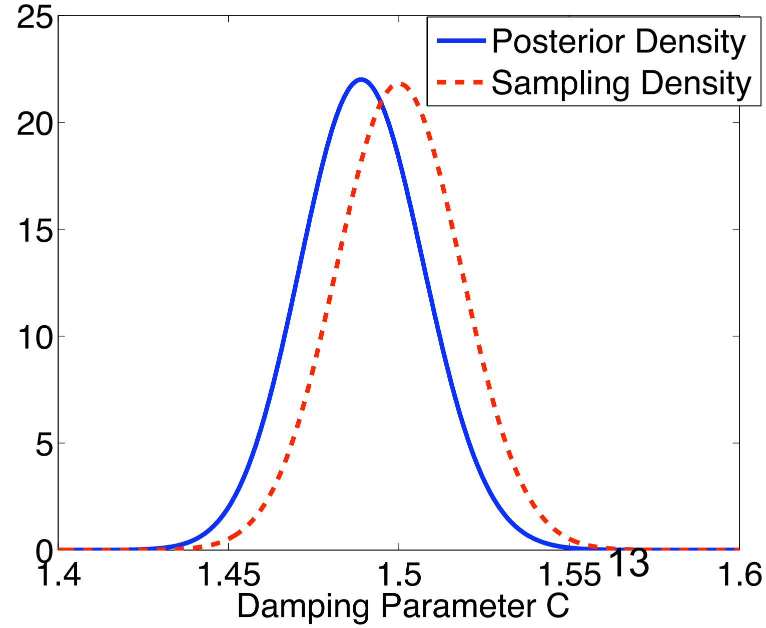
Midpoint formula:

$$\pi(\theta|y) \approx \frac{1}{\sum_{i=1}^k e^{-(SS_{\zeta_i} - SS_\theta)/2\sigma_0^2} w_i}$$

Note:

•Slow even for one parameter.

•Strategy: create Markov chain using random sampling so that created chain has the posterior distribution as its limiting (stationary) distribution.

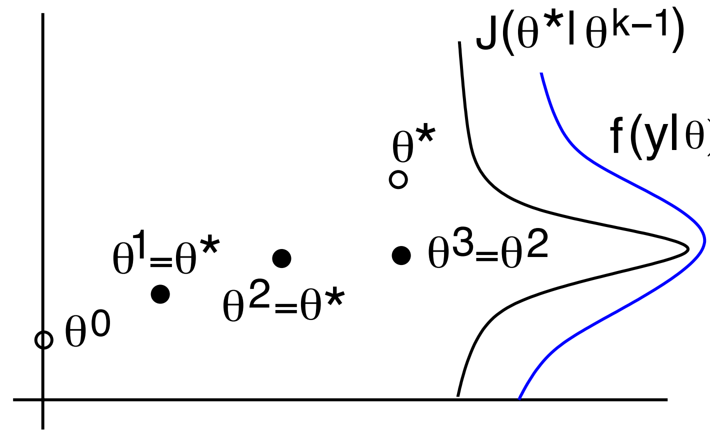


Bayesian Model Calibration

Bayesian Model Calibration:

•Parameters considered to be random variables with associated densities.

$$\pi(\theta|y) = \frac{f(y|\theta)\pi_0(\theta)}{\int_{\mathbb{R}^p} f(y|\theta)\pi_0(\theta)d\theta}$$



Problem:

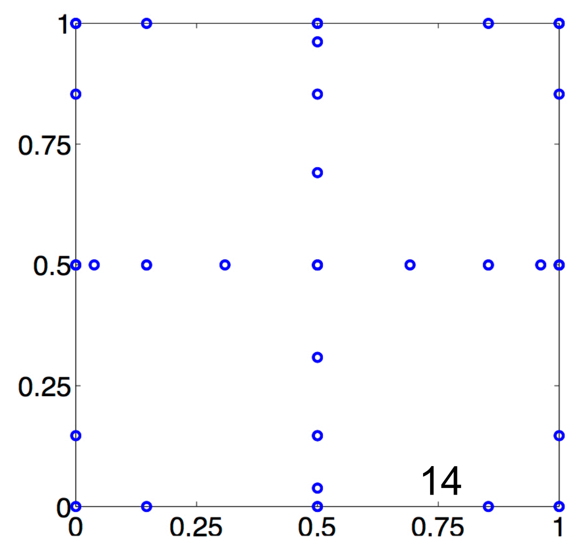
•Often requires high dimensional integration;

- ○ e.g., p = 18 for MFC model
- ○ p = thousands to millions for some models

Strategies:

•Sampling methods

•Sparse grid quadrature techniques



Markov Chains

Definition: Sequence of random variables $X_1, X_2, \cdots$ that satisfy Markov property:

X_{n+1} depends only on X_n; that is

$$P(X_{n+1} = x_{n+1} | X_0 = x_0, X_1 = x_1, \cdots, X_n = x_n) = P(X_{n+1} = x_{n+1} | X_n = x_n)$$

where x_i is the state of the chain at time i.

Note: A Markov chain is characterized by three components: a state space, an initial distribution, and a transition kernel.

State Space: Range of X_i: Set of all possible values

Initial Distribution: (Mass)

$$p^0 = [p_1^0, p_2^0, \cdots, p_n^0] \quad , \quad p_i^0 = P(X_0 = x_i)$$

Transition Probability: (Markov Kernel)

$$p_{ij} = P(X_{n+1} = x_j | X_n = x_i)$$

$$p_{ij}^{(n)} = P(X_{m+n} = x_j | X_m = x_i) \quad \text{(}n\text{-step transition probability)}$$

$$P = [p_{ij}] \quad , \quad P_n = [p_{ij}^{(n)}]$$

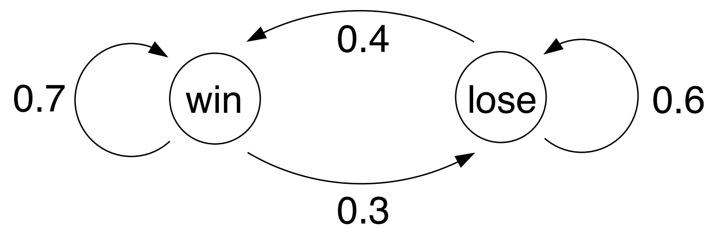
Markov Chain Techniques

Markov Chain: Sequence of events where current state depends only on last value.

Baseball: States are S = $\{$win,lose$\}$. Initial state is $p^0 = [0.8, 0.2]$.

•Assume that team which won last game has 70% chance of winning next game and 30% chance of losing next game.

•Assume losing team wins 40% and loses 60% of next games.



•Percentage of teams who win/lose next game given by

$$p^1 = [0.8\ ,\ 0.2] \begin{bmatrix} 0.7 & 0.3 \\ 0.4 & 0.6 \end{bmatrix} = [0.64\ ,\ 0.36]$$

•Question: does the following limit exist?

$$p^n = [0.8\ ,\ 0.2] \begin{bmatrix} 0.7 & 0.3 \\ 0.4 & 0.6 \end{bmatrix}^n$$

Markov Chain Techniques

Baseball Example: Solve constrained relation

$$\pi = \pi P \quad , \quad \sum \pi_i = 1$$

$$\Rightarrow [\pi_{win} \, , \, \pi_{lose}] \begin{bmatrix} 0.7 & 0.3 \\ 0.4 & 0.6 \end{bmatrix} = [\pi_{win} \, , \, \pi_{lose}] \quad , \quad \pi_{win} + \pi_{lose} = 1$$

to obtain

$$\pi = [0.5714 \, , \, 0.4286]$$

Markov Chain Techniques

Baseball Example: Solve constrained relation

$$\pi = \pi P \quad , \quad \sum \pi_i = 1$$

$$\Rightarrow [\pi_{win} \, , \, \pi_{lose}] \begin{bmatrix} 0.7 & 0.3 \\ 0.4 & 0.6 \end{bmatrix} = [\pi_{win} \, , \, \pi_{lose}] \quad , \quad \pi_{win} + \pi_{lose} = 1$$

to obtain

$$\pi = [0.5714 \, , \, 0.4286]$$

Alternative: Iterate to compute solution

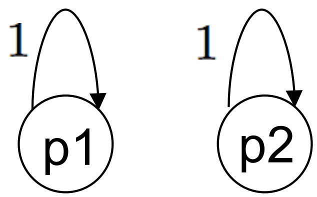
n	p^n	n	p^n	n	p^n
0	$[0.8000 \, , \, 0.2000]$	4	$[0.5733 \, , \, 0.4267]$	8	$[0.5714 \, , \, 0.4286]$
1	$[0.6400 \, , \, 0.3600]$	5	$[0.5720 \, , \, 0.4280]$	9	$[0.5714 \, , \, 0.4286]$
2	$[0.5920 \, , \, 0.4080]$	6	$[0.5716 \, , \, 0.4284]$	10	$[0.5714 \, , \, 0.4286]$
3	$[0.5776 \, , \, 0.4224]$	7	$[0.5715 \, , \, 0.4285]$		

Notes:

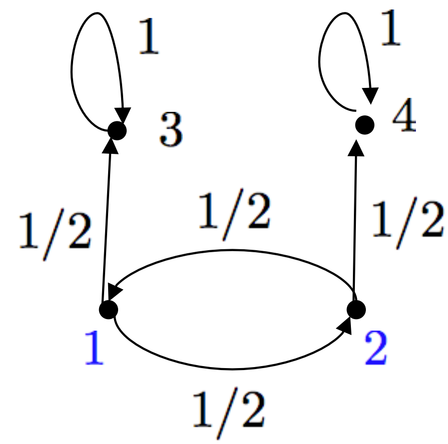
- Forms basis for Markov Chain Monte Carlo (MCMC) techniques
- Goal: construct chains whose stationary distribution is the posterior density

Irreducible Markov Chains

Reducible Markov Chain:



$$p^0 = [p1 \ , \ p2] = \pi$$



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

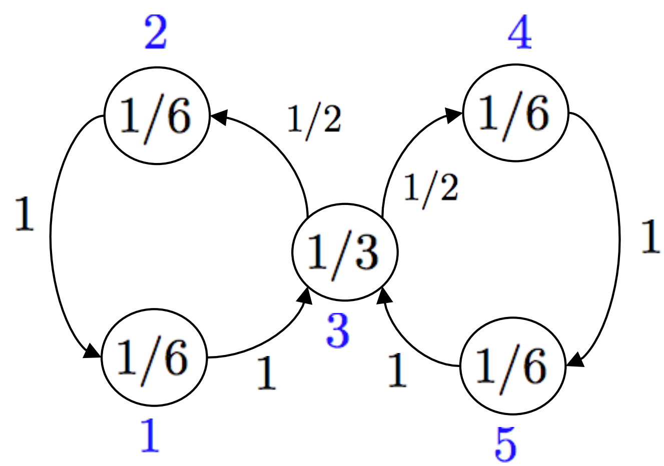
Note: Limiting distribution not unique if chain is reducible.

Irreducible: A Markov chain is *irreducible* if any state x_j and be reached from any state x_i in a finite number of steps; that is

$p_{ij}^{(n)} > 0$ for all states in finite n

Periodic Markov Chains

Example:



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

$$\pi = \begin{bmatrix} \frac{1}{6} & \frac{1}{6} & \frac{1}{3} & \frac{1}{6} & \frac{1}{6} \end{bmatrix}$$

Note: Chain returns to state 1 at steps $3, 6, 9, \cdots$ so Period = 3

Note: Probability mass "cycles" through chain so no convergence

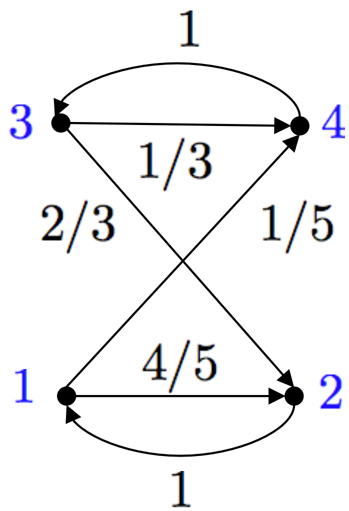
Periodicity: A Markov chain is *periodic* if parts of the state space are visited at regular intervals. The period *k* is defined as

$$\begin{aligned} k &= \gcd\left\{n | p_{ii}^{(n)} > 0\right\} \\ &= \gcd\left\{n | P(X_{m+n} = x_i | X_m = x_i) > 0\right\} \end{aligned}$$

- The chain is aperiodic if $k = 1$.

Periodic Markov Chains

Example:



$$P = \begin{bmatrix} 0 & 4/5 & 0 & 1/5 \\ 1 & 0 & 0 & 0 \\ 0 & 2/3 & 0 & 1/3 \\ 0 & 0 & 1 & 0 \end{bmatrix}$$

$$p^0 = \left[\begin{array}{cccc} \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} \end{array} \right]$$

$$p^0 = \left[\begin{array}{cccc} 1 & 0 & 0 & 0 \end{array} \right]$$

Stationary Distribution

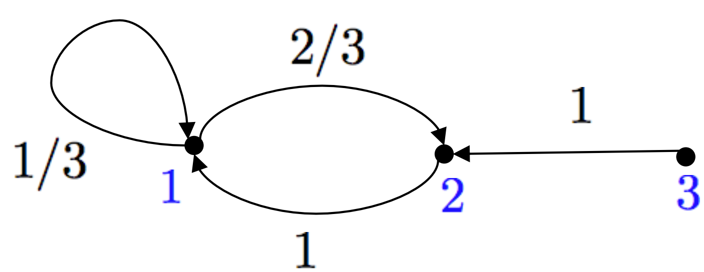
Theorem: A finite, homogeneous Markov chain that is irreducible and aperiodic has a unique stationary distribution π and the chain will converge in the sense of distributions from any initial distribution p^0.

Recurrence (Persistence): A state x_i is recurrent (persistent) if the probability of returning to x_i is 1; that is,

$$P(X_{m+n} = x_i \text{ for some } n \geq 1 | X_m = x_i) = 1$$

- It is *transient* if probability strictly less than 1

Example: State 3 is transient



Ergodicity: A state is termed *ergodic* if it is aperiodic and recurrent. If all states of an irreducible Markov chain are ergodic, the chain is said to be *ergodic*.

Matrix Theory

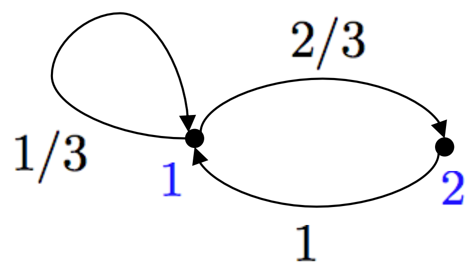
Definition: A matrix $A \in \mathbb{R}^{(n \times n)}$ is

(i) Nonnegative, denoted $A \geq 0$, if $a_{ij} \geq 0$ for all i, j

(ii) Strictly positive, denoted $A > 0$, if $a_{ij} > 0$ for all i, j

Lemma: Let P be the transition matrix of an ergodic finite Markov chain with state space S. Then for some $N_0 \geq 1$, $P_n > 0$ for all $n > N_0$.

Example:



$$P = \begin{bmatrix} 1/3 & 2/3 \\ 1 & 0 \end{bmatrix}$$

$$P_2 = \begin{bmatrix} 7/9 & 2/9 \\ 1/3 & 2/3 \end{bmatrix}$$

Matrix Theory

Theorem (Perron-Frobenius): For any strictly positive matrix $A > 0$, there exist $\lambda_0 > 0$ and $x_0 > 0$ such that

(i) $Ax_0 = \lambda_0 x_0$

(ii) If $\lambda \neq \lambda_0$ is any other eigenvalue of A, then $|\lambda| < \lambda_0$

(iii) λ_0 has geometric and algebraic multiplicity 1

Corollary 1: If $A \geq 0$ is a nonnegative matrix such that $A^n > 0$, then theorem also applies to A.

Proposition: Let $A > 0$ be a strictly positive $n \times n$ matrix with row and column sums

$$r_i = \sum_j a_{ij} \quad , \quad c_j = \sum_i a_{ij} \ , \ i,j = 1, \cdots, n$$

Then

$$\min_i r_i \leq \lambda_0 \leq \max_i r_i \quad , \quad \min_j c_j \leq \lambda_0 \leq \max_j c_j$$

Stationary Distribution

Corollary: Let $P \geq 0$ be the transition matrix of an ergodic Markov chain. Then there exists a unique stationary distribution π such that $\pi P = \pi$.

Proof: By Lemma and Corollary 1, P has a largest eigenvalue $\lambda_0 = 1$.

Since multiplicity is 1, unique π such that $\pi P = \pi$ and $\sum_i \pi_i = 1$.

Convergence: Express

$$UPV = \Lambda = \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & & \\ \vdots & & \ddots & \vdots \\ 0 & & \cdots & \lambda_k \end{bmatrix}$$

where $1 > |\lambda_2| \geqslant \cdots \geqslant |\lambda_k|$ and $V = U^{-1}$

Note:

$$\lim_{n\to\infty} P^n = \lim_{n\to\infty} V \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & \lambda_2^n & & \\ \vdots & & \ddots & \vdots \\ 0 & & \cdots & \lambda_k^n \end{bmatrix} U = V \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & 0 & & \\ \vdots & & \ddots & \vdots \\ 0 & & \cdots & 0 \end{bmatrix} U$$

Stationary Distribution

Note: $UP = \Lambda U$ implies

$$\begin{bmatrix} \pi_1 & \cdots & \pi_k \\ \vdots & & \vdots \\ u_{k1} & \cdots & u_{kk} \end{bmatrix} \begin{bmatrix} & P & \end{bmatrix} = \begin{bmatrix} 1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_n \end{bmatrix} \begin{bmatrix} \pi_1 & \cdots & \pi_k \\ \vdots & & \vdots \\ u_{k1} & \cdots & u_{kk} \end{bmatrix}$$

and $V = U^{-1} \Rightarrow$

$$UV = \begin{bmatrix} \pi_1 & \cdots & \pi_k \\ \vdots & & \vdots \\ u_{k1} & \cdots & u_{kk} \end{bmatrix} \begin{bmatrix} 1 & \cdots & v_{1k} \\ \vdots & & \vdots \\ 1 & \cdots & v_{kk} \end{bmatrix} = \begin{bmatrix} 1 & \cdots & 0 \\ \vdots & & \vdots \\ 0 & \cdots & 1 \end{bmatrix}$$

Thus

$$\begin{aligned} \lim_{n\to\infty} p^n &= \lim_{n\to\infty} p^0 P^n \\ &= \lim_{n\to\infty} [p_1^0, \dots, p_k^0] \begin{bmatrix} 1 & \cdots & v_{k1} \\ \vdots & & \vdots \\ 1 & \cdots & v_{kk} \end{bmatrix} \begin{bmatrix} 1 & & & \\ & \lambda_2^n & & \\ & & \ddots & \\ & & & \lambda_k^n \end{bmatrix} \begin{bmatrix} \pi_1 & \cdots & \pi_k \\ \vdots & & \vdots \\ u_{k1} & \cdots & u_{kk} \end{bmatrix} \\ &= \begin{bmatrix} p_1^0 & \cdots & p_k^0 \end{bmatrix} \begin{bmatrix} 1 & \cdots & v_{k1} \\ \vdots & & \vdots \\ 1 & \cdots & v_{kk} \end{bmatrix} \begin{bmatrix} 1 & & & \\ & 0 & & \\ & & \ddots & \\ & & & 0 \end{bmatrix} \begin{bmatrix} \pi_1 & \cdots & \pi_k \\ \vdots & & \vdots \\ u_{k1} & \cdots & u_{kk} \end{bmatrix} \\ &= [\pi_1, \dots, \pi_k] \\ &= \pi, \end{aligned}$$

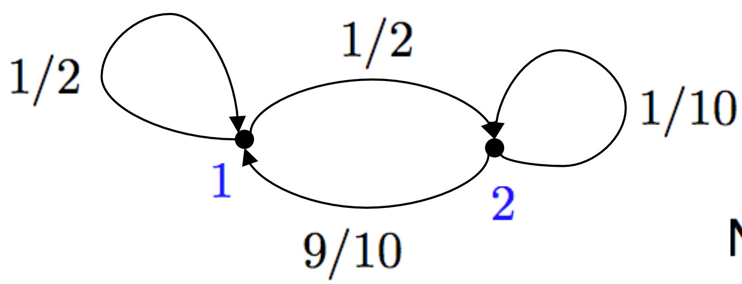
Detailed Balance Conditions

Reversible Chains: A Markov chain determined by the transition matrix $P = [p_{ij}]$ is reversible if there is a distribution π that satisfies the detailed balance conditions

$$\pi_i p_{ij} = \pi_j p_{ji}$$

Proof: We need to show that $\pi_j = \sum_i \pi_i p_{ij}$. Note that $\sum_i \pi_i p_{ij} = \sum_i \pi_j p_{ji} = \pi_j \sum_i p_{ji}$

Example:



$$P = \begin{bmatrix} 1/2 & 1/2 \\ 9/10 & 1/10 \end{bmatrix}$$

$$\pi = \begin{bmatrix} 9/14 & 5/14 \end{bmatrix}$$

Note: $\frac{1}{2} \cdot \frac{9}{14} = \frac{9}{10} \cdot \frac{5}{14}$ so detailed balance satisfied

Markov Chain Monte Carlo Methods

Strategy: Markov chain simulation used when it is impossible, or computationally prohibitive, to sample q directly from

$$\pi(\theta|y) = \frac{f(y|\theta)\pi_0(\theta)}{\int_{\mathbb{R}^p} f(y|\theta)\pi_0(\theta)d\theta}$$

- Create a Markov process whose stationary distribution is $\pi(\theta|y)$

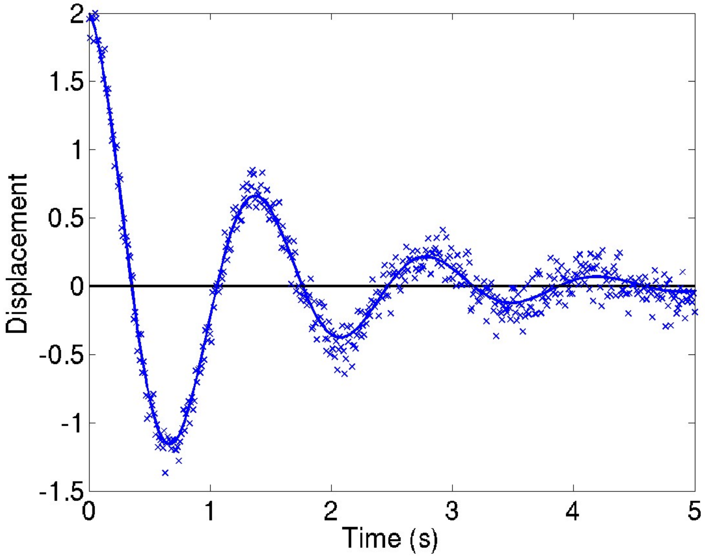
Note:

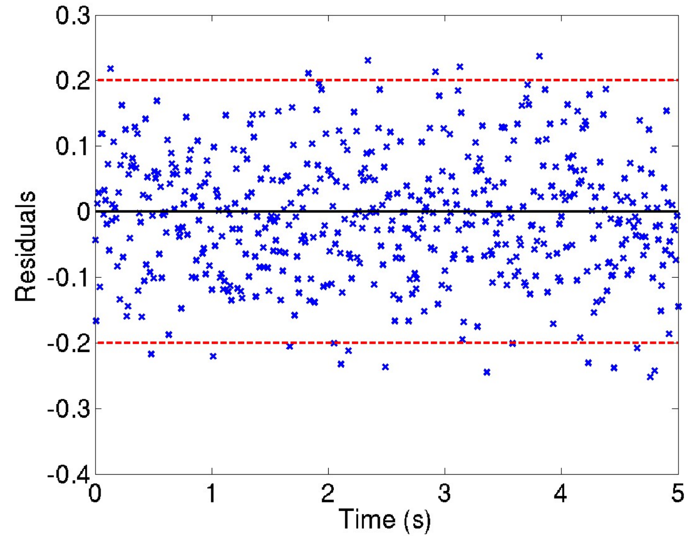
• In Markov chain theory, we are given a Markov chain, P, and we construct its equilibrium distribution.

• In MCMC theory, we are "given" a distribution and we want to construct a Markov chain that is reversible with respect to it.

Model Calibration Problem

Assumption: Assume that measurement errors are iid and $\varepsilon_i \sim N(0, \sigma^2)$





Likelihood:

$$f(y|\theta) = L(\theta, \sigma|y) = \frac{1}{(2\pi\sigma^2)^{n/2}} e^{-SS_\theta/2\sigma^2}$$

where

$$SS_\theta = \sum_{j=1}^{n} [y_j - f_i(\theta)]^2$$

is the sum of squares error.

Markov Chain Monte Carlo Methods

General Strategy:

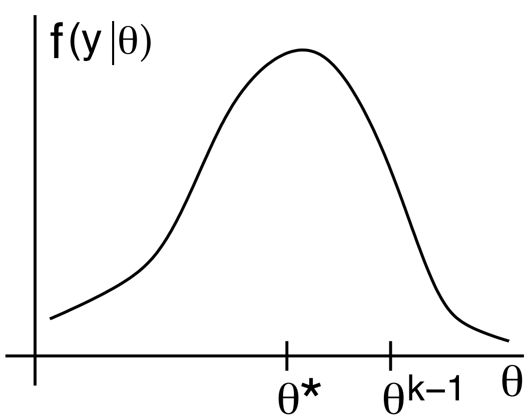
- Current value: $X_{k-1} = \theta^{k-1}$
- Propose candidate $\theta^* \sim J(\theta^*|\theta^{k-1})$ from proposal (jumping) distribution
- With probability $\alpha(\theta^*, \theta^{k-1})$, accept θ^*; i.e., $X_k = \theta^*$
- Otherwise, stay where you are: $X_k = \theta^{k-1}$

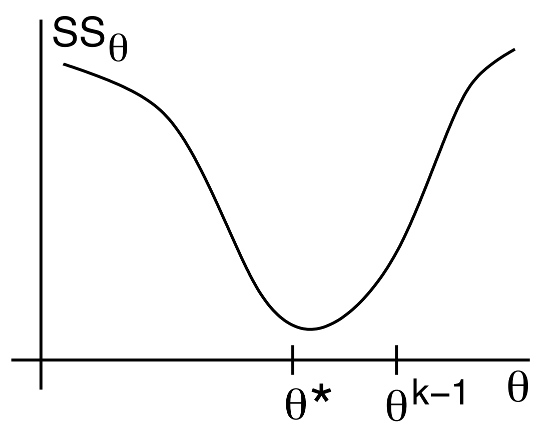
Intuition: Recall that

$$\pi(\theta|y) = \frac{f(y|\theta)\pi_0(\theta)}{\int_{\mathbb{R}^p} f(y|\theta)\pi_0(\theta)d\theta}$$

where

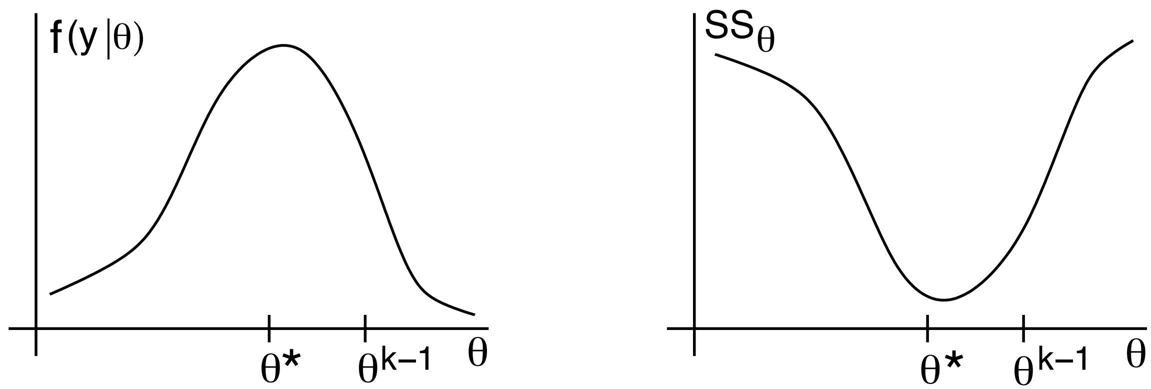
$$f(y|\theta) = \frac{1}{(2\pi\sigma^2)^{n/2}} e^{-\sum_{i=1}^n [y_i - f_i(\theta)]^2/2\sigma^2} = \frac{1}{(2\pi\sigma^2)^{n/2}} e^{-SS_\theta/2\sigma^2}$$





Markov Chain Monte Carlo Methods

Intuition:

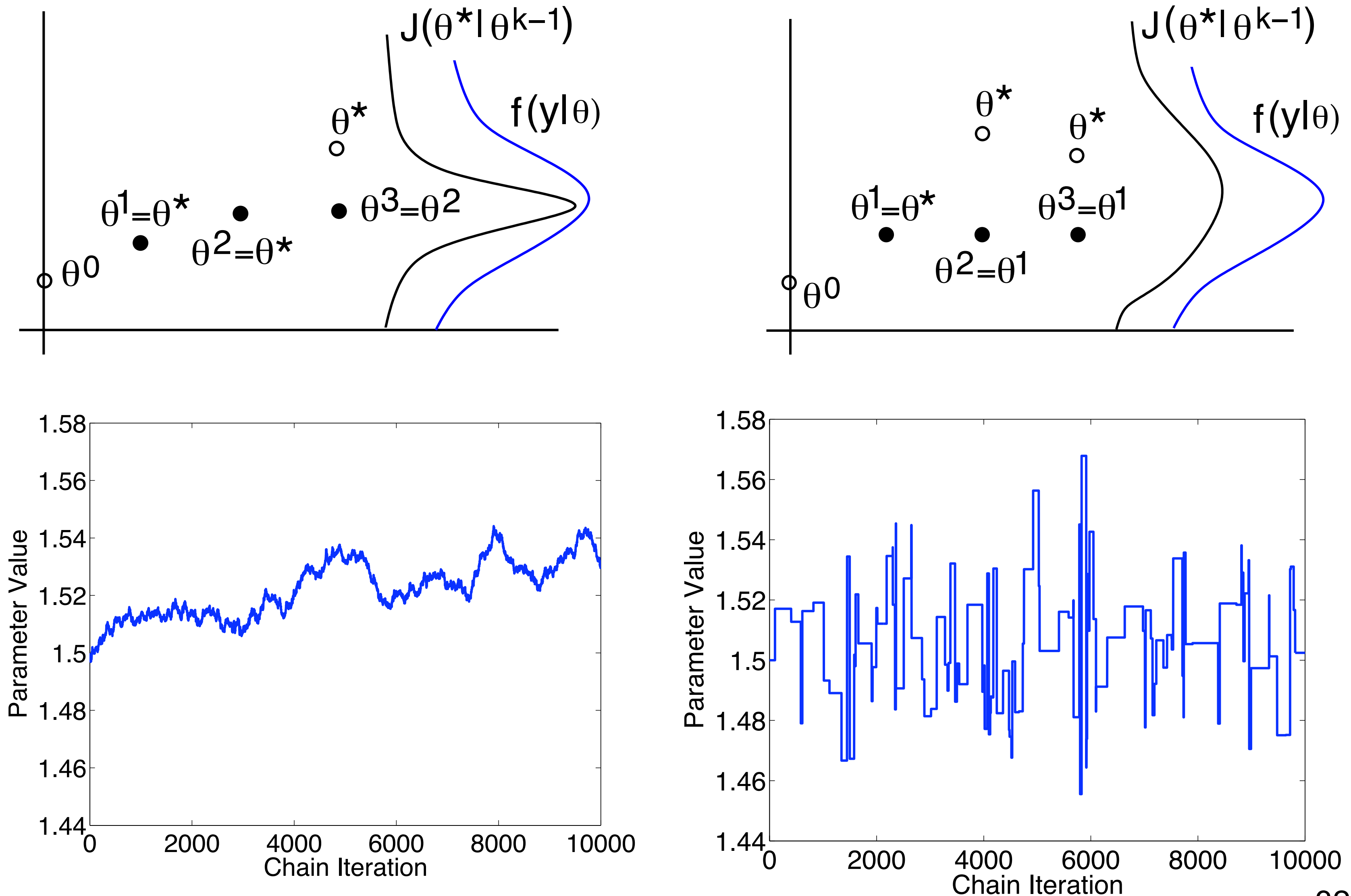


- Consider $r(\theta^*|\theta^{k-1}) = \frac{\pi(\theta^*|y)}{\pi(\theta^{k-1}|y)} = \frac{f(y|\theta^*)\pi_0(\theta^*)}{f(y|\theta^{k-1})\pi_0(\theta^{k-1})}$
  - If $r < 1 \Rightarrow f(y|\theta^*) < f(y|\theta^{k-1})$, accept with probability $\alpha = r$
  - If $r > 1$, accept with probability $\alpha = 1$

Note: Narrower proposal distribution yields higher probability of acceptance.

Markov Chain Monte Carlo Methods

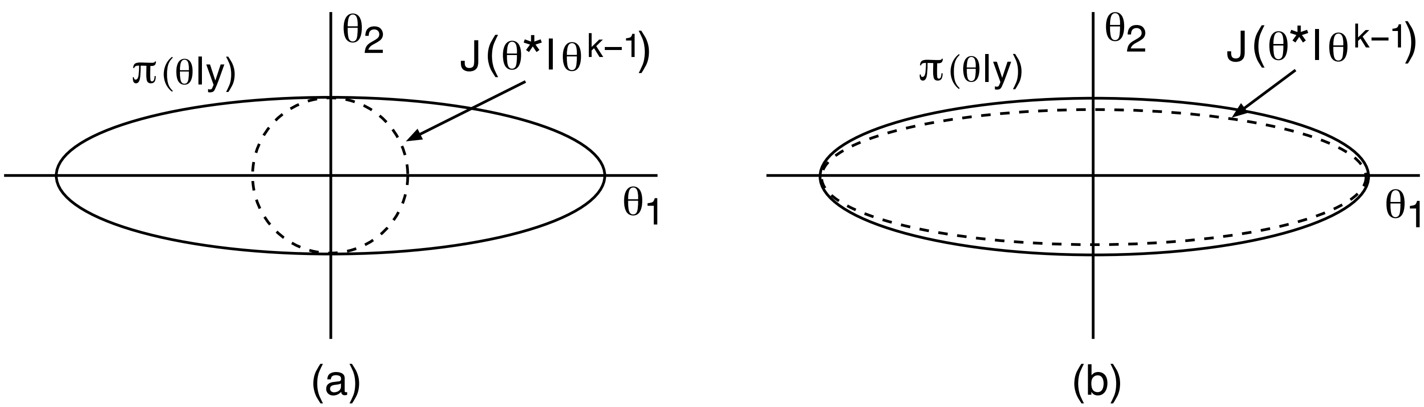
Note: Narrower proposal distribution yields higher probability of acceptance.



Proposal Distribution

Proposal Distribution: Significantly affects mixing

- Too wide: Too many points rejected and chain stays still for long periods;
- Too narrow: Acceptance ratio is high but algorithm is slow to explore parameter space
- Ideally, it should have similar "shape" to posterior distribution.



Problem:

- Anisotropic posterior, isotropic proposal;
- Efficiency nonuniform for different parameters

Result:

- Recovers efficiency of univariate case