Reversible Jump Markov Chain Monte Carlo

Based on Chapter 3 in Handbook of Markov Chain Monte Carlo

Yanan Fan   Scott A. Sisson

Talk by Nir Levin, July 2015

Outline

1. Introduction
   - Motivation
   - Model
   - The M.H Algorithm
   - M.H to Reversible Jump
   - Some Use Cases
2. Implementation
   - General Approach
   - Marginalization
   - Centering & Order Methods
   - Generic Samplers
3. Simulation
   - Model Formulation
   - Results

Outline

Start With The End

- Model is given by,

$$y = \sum_{i=1}^{n} x_i b_i + \varepsilon$$

- $n \leq m$ where m is the number of possible inputs, x_i's are the inputs, y is the output (what we measure), ε is the measurment noise.
- We wish to estimate $\{n, b_1, .., b_n\}$ using Bayesian modeling, given k independent measurments.
- Next, we define relevant distributions for the model.

Start With The End

- Model is given by,

$$y = \sum_{i=1}^{n} x_i b_i + \varepsilon$$

- $n \leq m$ where m is the number of possible inputs, x_i's are the inputs, y is the output (what we measure), ε is the measurment noise.
- We wish to estimate $\{n, b_1, .., b_n\}$ using Bayesian modeling, given k independent measurments.
- Next, we define relevant distributions for the model.

Start With The End

- Model is given by,

$$y = \sum_{i=1}^{n} x_i b_i + \varepsilon$$

- $n \leq m$ where m is the number of possible inputs, x_i's are the inputs, y is the output (what we measure), ε is the measurment noise.
- We wish to estimate $\{n, b_1, .., b_n\}$ using Bayesian modeling, given k independent measurments.
- Next, we define relevant distributions for the model.

Start With The End

- Model is given by,

$$y = \sum_{i=1}^{n} x_i b_i + \varepsilon$$

- $n \leq m$ where m is the number of possible inputs, x_i's are the inputs, y is the output (what we measure), ε is the measurment noise.
- We wish to estimate $\{n, b_1, .., b_n\}$ using Bayesian modeling, given k independent measurments.
- Next, we define relevant distributions for the model.

Starting With The End (Cont'd)

- The likelihood is given by,

$$L(y|n,b) = \left(\frac{1}{\sqrt{2\pi\sigma_0^2}}\right)^k \cdot \exp\left\{-\frac{\left\|y - X^T b\right\|^2}{2\sigma_0^2}\right\}$$

- The prior over b is given by,

$$p(b|n) = \left(\frac{1}{\sqrt{2\pi\sigma_p^2}}\right)^n \cdot \exp\left\{-\frac{\left\|b-\mu\right\|^2}{2\sigma_p^2}\right\}$$

- The prior over n is given by,

$$p(n) = \frac{1}{m}$$

Starting With The End (Cont'd)

- The likelihood is given by,

$$L(y|n,b) = \left(\frac{1}{\sqrt{2\pi\sigma_0^2}}\right)^k \cdot \exp\left\{-\frac{\|y - X^T b\|^2}{2\sigma_0^2}\right\}$$

- The prior over b is given by,

$$p(b|n) = \left(\frac{1}{\sqrt{2\pi\sigma_p^2}}\right)^n \cdot \exp\left\{-\frac{\|b - \mu\|^2}{2\sigma_p^2}\right\}$$

- The prior over n is given by,

$$p(n) = \frac{1}{m}$$

Starting With The End (Cont'd)

- The likelihood is given by,

$$L(y|n,b) = \left(\frac{1}{\sqrt{2\pi\sigma_0^2}}\right)^k \cdot \exp\left\{-\frac{\|y - X^T b\|^2}{2\sigma_0^2}\right\}$$

- The prior over b is given by,

$$p(b|n) = \left(\frac{1}{\sqrt{2\pi\sigma_p^2}}\right)^n \cdot \exp\left\{-\frac{\|b - \mu\|^2}{2\sigma_p^2}\right\}$$

- The prior over n is given by,

$$p(n) = \frac{1}{m}$$

Outline

Understanding the Problem

- Bayesian modeling context for observed data.
- Countable collection Models $\mathscr{M} = \{\mathscr{M}_1, \mathscr{M}_2, ..\}$ indexed by $k \in \mathscr{K}$.
- Each model $\mathscr{M}_k$ has an n_k-dimensional vector of unknown parameters, $\theta_k \in \mathbb{R}^{n_k}$. n_k can be different for different $k \in \mathscr{K}$.

Understanding the Problem

- Bayesian modeling context for observed data.
- Countable collection Models $\mathscr{M} = \{\mathscr{M}_1, \mathscr{M}_2, ..\}$ indexed by $k \in \mathscr{K}$.
- Each model $\mathscr{M}_k$ has an n_k-dimensional vector of unknown parameters, $\theta_k \in \mathbb{R}^{n_k}$. n_k can be different for different $k \in \mathscr{K}$.

Understanding the Problem

- Bayesian modeling context for observed data.
- Countable collection Models $\mathscr{M} = \{\mathscr{M}_1, \mathscr{M}_2, ..\}$ indexed by $k \in \mathscr{K}$.
- Each model $\mathscr{M}_k$ has an n_k-dimensional vector of unknown parameters, $\theta_k \in \mathbb{R}^{n_k}$. n_k can be different for different $k \in \mathscr{K}$.

Understanding the Problem (Cont'd)

- The joint posterior of $(k, \theta_k|x)$ is given by,

$$\pi(k, \theta_k|x) = \frac{L(x|k, \theta_k)\, p(\theta_k|k)\, p(k)}{\sum_{k' \in \mathscr{K}} \int_{\mathbb{R}^{n_{k'}}} L\left(x|k', \theta'_{k'}\right) p\left(\theta'_{k'}|k'\right) p(k')}$$

- $L(x|k, \theta_k)$ is the likelihood distribution, $p(\theta_k|k)$ is the prior of θ_k under model $\mathscr{M}_k$ and $p(k)$ is the prior for model $\mathscr{M}_k$.
- Reversible Jump sets the posterior distribution as the target of MCMC sampler over the state space $\Theta = \cup_{k \in \mathscr{K}} \left(\{k\} \times \mathbb{R}^{n_k}\right)$.

Understanding the Problem (Cont'd)

- The joint posterior of $(k, \theta_k | x)$ is given by,

$$\pi(k, \theta_k | x) = \frac{L(x|k, \theta_k)\, p(\theta_k|k)\, p(k)}{\sum_{k' \in \mathscr{K}} \int_{\mathbb{R}^{n_{k'}}} L\left(x|k', \theta'_{k'}\right) p\left(\theta'_{k'}|k'\right) p(k')}$$

- $L(x|k, \theta_k)$ is the likelihood distribution, $p(\theta_k|k)$ is the prior of θ_k under model $\mathscr{M}_k$ and $p(k)$ is the prior for model $\mathscr{M}_k$.
- Reversible Jump sets the posterior distribution as the target of MCMC sampler over the state space $\Theta = \cup_{k \in \mathscr{K}} \left(\{k\} \times \mathbb{R}^{n_k}\right)$.

Understanding the Problem (Cont'd)

- The joint posterior of $(k, \theta_k | x)$ is given by,

$$\pi(k, \theta_k | x) = \frac{L(x|k, \theta_k)\, p(\theta_k|k)\, p(k)}{\sum_{k' \in \mathscr{K}} \int_{\mathbb{R}^{n_{k'}}} L\left(x|k', \theta'_{k'}\right) p\left(\theta'_{k'}|k'\right) p(k')}$$

- $L(x|k, \theta_k)$ is the likelihood distribution, $p(\theta_k|k)$ is the prior of θ_k under model $\mathscr{M}_k$ and $p(k)$ is the prior for model $\mathscr{M}_k$.
- Reversible Jump sets the posterior distribution as the target of MCMC sampler over the state space $\Theta = \cup_{k \in \mathscr{K}} (\{k\} \times \mathbb{R}^{n_k})$.

Outline

Reminder

- Propose some transition distribution $q(\theta, \theta') \equiv q(\theta'|\theta)$ and target distribution π.
- Starting with some initial state θ_0, the algorithm proceeds as follows:

  1. Given the current state θ, generate $\theta' \sim q(\theta, \cdot)$.
  2. Set

$$\theta := \begin{cases} \theta' & \text{w.p. } \alpha(\theta, \theta') \\ \theta & \text{otherwise} \end{cases}$$

  where

$$\alpha = \min\left\{1, \frac{\pi(\theta')\, q(\theta', \theta)}{\pi(\theta)\, q(\theta, \theta')}\right\}$$

Reminder

- Propose some transition distribution $q(\theta, \theta') \equiv q(\theta'|\theta)$ and target distribution π.
- Starting with some initial state θ_0, the algorithm proceeds as follows:

1. Given the current state θ, generate $\theta' \sim q(\theta, \cdot)$.
2. Set

$$\theta := \begin{cases} \theta' & \text{w.p. } \alpha(\theta, \theta') \\ \theta & \text{otherwise} \end{cases}$$

where

$$\alpha = \min\left\{1, \frac{\pi(\theta')q(\theta', \theta)}{\pi(\theta)q(\theta, \theta')}\right\}$$

Reminder

- Propose some transition distribution $q(\theta, \theta') \equiv q(\theta'|\theta)$ and target distribution π.
- Starting with some initial state θ_0, the algorithm proceeds as follows:
  1. Given the current state θ, generate $\theta' \sim q(\theta, \cdot)$.
  2. Set
  $$\theta := \begin{cases} \theta' & \text{w.p. } \alpha(\theta, \theta') \\ \theta & \text{otherwise} \end{cases}$$
  where
  $$\alpha = \min\left\{1, \frac{\pi(\theta')q(\theta', \theta)}{\pi(\theta)q(\theta, \theta')}\right\}$$

Reminder

- Propose some transition distribution $q(\theta, \theta') \equiv q(\theta'|\theta)$ and target distribution π.
- Starting with some initial state θ_0, the algorithm proceeds as follows:

1. Given the current state θ, generate $\theta' \sim q(\theta, \cdot)$.
2. Set

$$\theta := \begin{cases} \theta' & \text{w.p. } \alpha(\theta, \theta') \\ \theta & \text{otherwise} \end{cases}$$

where

$$\alpha = \min\left\{1, \frac{\pi(\theta')q(\theta', \theta)}{\pi(\theta)q(\theta, \theta')}\right\}$$

Note that..

- The motivation behind M.H algorithm is satisfying the detailed balance condition, so that π is the stationary distribution.
- Following the same concept we can suggest MCMC for multi-models as well, let's call it Reversible Jump MCMC.

Note that..

- The motivation behind M.H algorithm is satisfying the detailed balance condition, so that π is the stationary distribution.
- Following the same concept we can suggest MCMC for multi-models as well, let's call it Reversible Jump MCMC.

Outline

Basically the Same

- Initialize the state for some (k, θ_k).
- Proceed as follows:
  - Generate a new state $\left(k', \theta'_{k'}\right)$ by,
    1. *Within-model move*: fixed k, update $\tilde{\theta}_k$.
    2. *Between-models move*: generate $\left(k', \theta'_{k'}\right)$ using $\left(k, \tilde{\theta}_k\right)$ and a mapping function.
  - Use acceptance/rejection scheme to determine next state.

Basically the Same

- Initialize the state for some (k, θ_k).
- Proceed as follows:
  - Generate a new state $\left(k', \theta'_{k'}\right)$ by,
    1. *Within-model move*: fixed k, update $\tilde{\theta}_k$.
    2. *Between-models move*: generate $\left(k', \theta'_{k'}\right)$ using $\left(k, \tilde{\theta}_k\right)$ and a mapping function.
  - Use acceptance/rejection scheme to determine next state.

Basically the Same

- Initialize the state for some (k, θ_k).
- Proceed as follows:
  - Generate a new state $\left(k', \theta'_{k'}\right)$ by,
    1. *Within-model move*: fixed k, update $\tilde{\theta}_k$.
    2. *Between-models move*: generate $\left(k', \theta'_{k'}\right)$ using $\left(k, \tilde{\theta}_k\right)$ and a mapping function.
  - Use acceptance/rejection scheme to determine next state.

Basically the Same

- Initialize the state for some (k, θ_k).
- Proceed as follows:
  - Generate a new state $\left(k', \theta'_{k'}\right)$ by,
    1. *Within-model move*: fixed k, update $\tilde{\theta}_k$.
    2. *Between-models move*: generate $\left(k', \theta'_{k'}\right)$ using $\left(k, \tilde{\theta}_k\right)$ and a mapping function.
  - Use acceptance/rejection scheme to determine next state.

Outline

Examples Time

Finite Mixture Models

Model is in the form

$$f(x|\theta_k) = \sum_{j=1}^{k} \omega_j f_j(x|\phi_j)$$

where $\theta_k = (\phi_1, .., \phi_k)$, $\sum_{j=1}^{k} \omega_j = 1$, and the number of mixture models, k, is also unknown.

Examples Time (Cont'd)

Variable Selection

Model is in the form

$$Y = X_\gamma \beta_\gamma + \varepsilon$$

where γ is a binary vector indexing the subset of X to be included, β are the regression coefficients and ε is noise.

Examples Time (Cont'd)

Autoregressive Process

Model is in the form

$$X_t = \sum_{\tau=1}^{k} a_\tau X_{t-\tau} + \varepsilon_t$$

where a_τ are the coefficients, ε_t is white noise and the order, k, is also unknown.

Outline

In Practice - "Dimension Matching"

- Say we wish to perform transition to a model with higher dimension, $\mathscr{M}_k : (k, \theta_k) \rightarrow \mathscr{M}_{k'} : \left(k', \theta'_{k'}\right)$.

1. Generate a RV u of length $d_{k \rightarrow k'} = n_{k'} - n_k$ from a known distribution $q_{d_{k \rightarrow k'}}(u)$.
2. Use a deterministic one-to-one mapping $\theta'_{k'} = g_{k \rightarrow k'}(\theta_k, u)$.
3. Accept new state w.p.

$$\alpha\left[(k, \theta_k), \left(k', \theta'_{k'}\right)\right] = \min\{1, A\}$$

$$A = \frac{\pi\left(k', \theta'_{k'}\right) q(k' \rightarrow k)}{\pi(k, \theta_k) q(k \rightarrow k') q_{d_{k \rightarrow k'}}(u)} \left|\frac{\partial g_{k \rightarrow k'}(\theta_k, u)}{\partial(\theta_k, u)}\right|$$

In Practice - "Dimension Matching"

- Say we wish to perform transition to a model with higher dimension, $\mathscr{M}_k : (k, \theta_k) \rightarrow \mathscr{M}_{k'} : \left(k', \theta^{'}_{k'}\right)$.

1. Generate a RV u of length $d_{k \rightarrow k'} = n_{k'} - n_k$ from a known distribution $q_{d_{k \rightarrow k'}}(u)$.
2. Use a deterministic one-to-one mapping $\theta^{'}_{k'} = g_{k \rightarrow k'}(\theta_k, u)$.
3. Accept new state w.p.

$$\alpha\left[(k, \theta_k), \left(k', \theta^{'}_{k'}\right)\right] = \min\{1, A\}$$

$$A = \frac{\pi\left(k', \theta^{'}_{k'}\right) q(k' \rightarrow k)}{\pi(k, \theta_k) q(k \rightarrow k') q_{d_{k \rightarrow k'}}(u)} \left|\frac{\partial g_{k \rightarrow k'}(\theta_k, u)}{\partial(\theta_k, u)}\right|$$

In Practice - "Dimension Matching"

- Say we wish to perform transition to a model with higher dimension, $\mathscr{M}_k : (k, \theta_k) \to \mathscr{M}_{k'} : \left(k', \theta'_{k'}\right)$.

1. Generate a RV u of length $d_{k\to k'} = n_{k'} - n_k$ from a known distribution $q_{d_{k\to k'}}(u)$.
2. Use a deterministic one-to-one mapping $\theta'_{k'} = g_{k\to k'}(\theta_k, u)$.
3. Accept new state w.p.

$$\alpha\left[(k, \theta_k), \left(k', \theta'_{k'}\right)\right] = \min\{1, A\}$$

$$A = \frac{\pi\left(k', \theta'_{k'}\right) q(k' \to k)}{\pi(k, \theta_k)\, q(k \to k')\, q_{d_{k\to k'}}(u)} \left|\frac{\partial g_{k\to k'}(\theta_k, u)}{\partial(\theta_k, u)}\right|$$

In Practice - "Dimension Matching"

- Say we wish to perform transition to a model with higher dimension, $\mathscr{M}_k : (k, \theta_k) \to \mathscr{M}_{k'} : \left(k', \theta^{'}_{k'}\right)$.

1. Generate a RV u of length $d_{k \to k'} = n_{k'} - n_k$ from a known distribution $q_{d_{k \to k'}}(u)$.
2. Use a deterministic one-to-one mapping $\theta^{'}_{k'} = g_{k \to k'}(\theta_k, u)$.
3. Accept new state w.p.

$$\alpha \left[(k, \theta_k), \left(k', \theta^{'}_{k'}\right) \right] = \min \{1, A\}$$

$$A = \frac{\pi\left(k', \theta^{'}_{k'}\right) q(k' \to k)}{\pi(k, \theta_k) q(k \to k') q_{d_{k \to k'}}(u)} \left| \frac{\partial g_{k \to k'}(\theta_k, u)}{\partial(\theta_k, u)} \right|$$

Some Notes

- The term $q(k \to k')$ denotes probability distribution over transition $\mathscr{M}_k \to \mathscr{M}_{k'}$.
- In this setup the transition from higher to lower dimension is deterministic.
- More generally, we can simply impose

$$n_k + d_{k \to k'} = n_{k'} + d_{k' \to k}$$

Some Notes

- The term $q(k \to k')$ denotes probability distribution over transition $\mathscr{M}_k \to \mathscr{M}_{k'}$.
- In this setup the transition from higher to lower dimension is deterministic.
- More generally, we can simply impose

$$n_k + d_{k \to k'} = n_{k'} + d_{k' \to k}$$

Some Notes

- The term $q(k \to k')$ denotes probability distribution over transition $\mathscr{M}_k \to \mathscr{M}_{k'}$.
- In this setup the transition from higher to lower dimension is deterministic.
- More generally, we can simply impose

$$n_k + d_{k \to k'} = n_{k'} + d_{k' \to k}$$

Example Time

- Two models,

$$\mathscr{M}_1 : (k = 1, \theta_1 \in \mathbb{R}), \quad \mathscr{M}_2 : \left(k = 2, \theta_2 \in \mathbb{R}^2\right)$$

- Transition $\mathscr{M}_1 \to \mathscr{M}_2$ (stochastic)

$$\theta_2^{(1)} = \theta_1 + u, \quad \theta_2^{(2)} = \theta_1 - u$$

- Transition $\mathscr{M}_2 \to \mathscr{M}_1$ (deterministic)

$$\theta_1 = \frac{1}{2}\left(\theta_2^{(1)} + \theta_2^{(2)}\right)$$

Example Time

- Two models,

$$\mathscr{M}_1 : (k = 1, \theta_1 \in \mathbb{R}), \quad \mathscr{M}_2 : \left(k = 2, \theta_2 \in \mathbb{R}^2\right)$$

- Transition $\mathscr{M}_1 \to \mathscr{M}_2$ (stochastic)

$$\theta_2^{(1)} = \theta_1 + u, \quad \theta_2^{(2)} = \theta_1 - u$$

- Transition $\mathscr{M}_2 \to \mathscr{M}_1$ (deterministic)

$$\theta_1 = \frac{1}{2}\left(\theta_2^{(1)} + \theta_2^{(2)}\right)$$

Example Time

- Two models,

$$\mathscr{M}_1 : (k = 1, \theta_1 \in \mathbb{R}), \quad \mathscr{M}_2 : \left(k = 2, \theta_2 \in \mathbb{R}^2\right)$$

- Transition $\mathscr{M}_1 \to \mathscr{M}_2$ (stochastic)

$$\theta_2^{(1)} = \theta_1 + u, \quad \theta_2^{(2)} = \theta_1 - u$$

- Transition $\mathscr{M}_2 \to \mathscr{M}_1$ (deterministic)

$$\theta_1 = \frac{1}{2}\left(\theta_2^{(1)} + \theta_2^{(2)}\right)$$

Outline

Making Life Easier

- In some Bayesian models, integrating over all (or parts of) θ_k is possible.
- Usually arises when assuming some structures such as conjugate priors.
- Results in smaller state space, thus often easier to implement.

Making Life Easier

- In some Bayesian models, integrating over all (or parts of) θ_k is possible.
- Usually arises when assuming some structures such as conjugate priors.
- Results in smaller state space, thus often easier to implement.

Making Life Easier

- In some Bayesian models, integrating over all (or parts of) θ_k is possible.
- Usually arises when assuming some structures such as conjugate priors.
- Results in smaller state space, thus often easier to implement.

Example Time

Revisit Variable Selection

Remeber the model,

$$Y = X_\gamma \beta_\gamma + \varepsilon$$

- Under some prior assumptions, β can be integrated out of the posterior.
- A Gibbs Sampler (for example) is available directly on γ.

Example Time

Revisit Variable Selection

Remeber the model,

$$Y = X_\gamma \beta_\gamma + \varepsilon$$

- Under some prior assumptions, β can be integrated out of the posterior.
- A Gibbs Sampler (for example) is available directly on γ.

Example Time

Revisit Variable Selection

Remeber the model,

$$Y = X_\gamma \beta_\gamma + \varepsilon$$

- Under some prior assumptions, β can be integrated out of the posterior.
- A Gibbs Sampler (for example) is available directly on γ.

Outline

Helpful Heuristics - Centering Methods

- A class of methods to achieve scaling based on "local" between-models moves. Assumes $g_{k \to k'}$ is given.
- Based on equal likelihood values under current and proposed models.
- Say we wish to move $\mathscr{M}_k \to \mathscr{M}_{k'}$. The RV "centering point" $c_{k \to k'}(\theta_k) = g_{k \to k'}(\theta_k, u)$, is u such that,

$$L(x|k, \theta_k) = L\left(x|k', c_{k \to k'}(\theta_k)\right)$$

- We can buillt more efficient transitions, based on likelihood.

Helpful Heuristics - Centering Methods

- A class of methods to achieve scaling based on "local" between-models moves. Assumes $g_{k \to k'}$ is given.
- Based on equal likelihood values under current and proposed models.
- Say we wish to move $\mathscr{M}_k \to \mathscr{M}_{k'}$. The RV "centering point" $c_{k \to k'}(\theta_k) = g_{k \to k'}(\theta_k, u)$, is u such that,

$$L(x|k, \theta_k) = L(x|k', c_{k \to k'}(\theta_k))$$

- We can buillt more efficient transitions, based on likelihood.

Helpful Heuristics - Centering Methods

- A class of methods to achieve scaling based on "local" between-models moves. Assumes $g_{k \to k'}$ is given.
- Based on equal likelihood values under current and proposed models.
- Say we wish to move $\mathcal{M}_k \to \mathcal{M}_{k'}$. The RV "centering point" $c_{k \to k'}(\theta_k) = g_{k \to k'}(\theta_k, u)$, is u such that,

$$L(x|k, \theta_k) = L\left(x|k', c_{k \to k'}(\theta_k)\right)$$

- We can buillt more efficient transitions, based on likelihood.

Helpful Heuristics - Centering Methods

- A class of methods to achieve scaling based on "local" between-models moves. Assumes $g_{k \to k'}$ is given.
- Based on equal likelihood values under current and proposed models.
- Say we wish to move $\mathscr{M}_k \to \mathscr{M}_{k'}$. The RV "centering point" $c_{k \to k'}(\theta_k) = g_{k \to k'}(\theta_k, u)$, is u such that,

$$L(x|k, \theta_k) = L\left(x|k', c_{k \to k'}(\theta_k)\right)$$

- We can buillt more efficient transitions, based on likelihood.

Example Time

Revisit Autoregressive Process

Remeber the model,

$$X_t = \sum_{\tau=1}^{k} a_\tau X_{t-\tau} + \varepsilon_t$$

- Say $\mathscr{M}_k$ is a model of order k, and $\mathscr{M}_{k'}$ is a model order $k+1$. We wish to move $\mathscr{M}_k \to \mathscr{M}_{k'}$.
- In this case, we have $c_{k\to k'}(\theta_k) = g_{k\to k'}(\theta_k, 0)$, since

$$L(x|k,\theta_k) = L\left(x|k',(\theta_k,0)\right)$$

Example Time

Revisit Autoregressive Process

Remeber the model,

$$X_t = \sum_{\tau=1}^{k} a_\tau X_{t-\tau} + \varepsilon_t$$

- Say $\mathscr{M}_k$ is a model of order k, and $\mathscr{M}_{k'}$ is a model order $k+1$. We wish to move $\mathscr{M}_k \to \mathscr{M}_{k'}$.
- In this case, we have $c_{k\to k'}(\theta_k) = g_{k\to k'}(\theta_k, 0)$, since

$$L(x|k,\theta_k) = L(x|k',(\theta_k,0))$$

Example Time

Revisit Autoregressive Process

Remeber the model,

$$X_t = \sum_{\tau=1}^{k} a_\tau X_{t-\tau} + \varepsilon_t$$

- Say $\mathscr{M}_k$ is a model of order k, and $\mathscr{M}_{k'}$ is a model order $k+1$. We wish to move $\mathscr{M}_k \to \mathscr{M}_{k'}$.
- In this case, we have $c_{k\to k'}(\theta_k) = g_{k\to k'}(\theta_k, 0)$, since

$$L(x|k,\theta_k) = L\left(x|k',(\theta_k,0)\right)$$

Helpful Heuristics - Order Methods

- Assume we have the "centering" constraint on u. We wish to determine a suitable $q_{k \to k'}(u)$.
- If u is a scalar, choose $q_{k \to k'}$ such that $\alpha[(k, \theta_k), (k', c_{k \to k'}(\theta_k))]$ is exactly one.
  Intuition: Proposals close to the "centering point" will also have large acceptance probability.
- If u is n-length vector, choose $q_{k \to k'}$ such that $\nabla^n \alpha[(k, \theta_k), (k', c_{k \to k'}(\theta_k))] = 0$.
  Intuition: Accepatance probability becomes flatter around $c_{k \to k'}(\theta_k)$.

Helpful Heuristics - Order Methods

- Assume we have the "centering" constraint on u. We wish to determine a suitable $q_{k\to k'}(u)$.
- If u is a scalar, choose $q_{k\to k'}$ such that $\alpha[(k,\theta_k),(k',c_{k\to k'}(\theta_k))]$ is exactly one.
  Intuition: Proposals close to the "centering point" will also have large acceptance probability.
- If u is n-length vector, choose $q_{k\to k'}$ such that $\nabla^n\alpha[(k,\theta_k),(k',c_{k\to k'}(\theta_k))]=0$.
  Intuition: Accepatance probability becomes flatter around $c_{k\to k'}(\theta_k)$.

Helpful Heuristics - Order Methods

- Assume we have the "centering" constraint on u. We wish to determine a suitable $q_{k\to k'}(u)$.
- If u is a scalar, choose $q_{k\to k'}$ such that $\alpha[(k,\theta_k),(k',c_{k\to k'}(\theta_k))]$ is exactly one.
  Intuition: Proposals close to the "centering point" will also have large acceptance probability.
- If u is n-length vector, choose $q_{k\to k'}$ such that $\nabla^n\alpha[(k,\theta_k),(k',c_{k\to k'}(\theta_k))]=0$.
  Intuition: Accepatance probability becomes flatter around $c_{k\to k'}(\theta_k)$.

Outline

Let's Make Everything Generic

- In centering & order methods, we assume we know $g_{k \to k'}$. But how one should choose it?
- Moreover, it's not necessarily easy to compute $q_{k \to k'}$ from $g_{k \to k'}$ using the above heuristics.
- New stradegy - move towards a more generic proposal machanism.
- Currently remains in the research horizon.

Let's Make Everything Generic

- In centering & order methods, we assume we know $g_{k \to k'}$. But how one should choose it?
- Moreover, it's not necessarily easy to compute $q_{k \to k'}$ from $g_{k \to k'}$ using the above heuristics.
- New stradegy - move towards a more generic proposal machanism.
- Currently remains in the research horizon.

Let's Make Everything Generic

- In centering & order methods, we assume we know $g_{k \to k'}$. But how one should choose it?
- Moreover, it's not necessarily easy to compute $q_{k \to k'}$ from $g_{k \to k'}$ using the above heuristics.
- New stradegy - move towards a more generic proposal machanism.
- Currently remains in the research horizon.

Let's Make Everything Generic

- In centering & order methods, we assume we know $g_{k \to k'}$. But how one should choose it?
- Moreover, it's not necessarily easy to compute $q_{k \to k'}$ from $g_{k \to k'}$ using the above heuristics.
- New stradegy - move towards a more generic proposal machanism.
- Currently remains in the research horizon.

Example Time

Green's Proposal

Analogy of Random-Walk Metropolis Sampler.

- Suppose that for each $k \in \mathscr{K}$, first and second moments of θ_k are available (denote by μ_k and $B_k B_k^T$ respectively).
- Suppose we wish to move from (k, θ_k) to model $\mathscr{M}_{k'}$.
- A new parameter is proposed by...

Example Time

Green's Proposal

Analogy of Random-Walk Metropolis Sampler.

- Suppose that for each $k \in \mathscr{K}$, first and second moments of θ_k are available (denote by μ_k and $B_k B_k^T$ respectively).
- Suppose we wish to move from (k, θ_k) to model $\mathscr{M}_{k'}$.
- A new parameter is proposed by...

Example Time

Green's Proposal

Analogy of Random-Walk Metropolis Sampler.

- Suppose that for each $k \in \mathscr{K}$, first and second moments of θ_k are available (denote by μ_k and $B_k B_k^T$ respectively).
- Suppose we wish to move from (k, θ_k) to model $\mathscr{M}_{k'}$.
- A new parameter is proposed by...

Example Time

Green's Proposal

Analogy of Random-Walk Metropolis Sampler.

- Suppose that for each $k \in \mathscr{K}$, first and second moments of θ_k are available (denote by μ_k and $B_k B_k^T$ respectively).
- Suppose we wish to move from (k, θ_k) to model $\mathscr{M}_{k'}$.
- A new parameter is proposed by...

Example Time (Cont'd)

Green's Proposal

$$\theta'_{k'} = \begin{cases} \mu_{k'} + B_{k'} \left[R B_k^{-1} (\theta_k - \mu_k) \right]_1^{n_{k'}} & \text{if } n_{k'} < n_k \\ \mu_{k'} + B_{k'} R B_k^{-1} (\theta_k - \mu_k) & \text{if } n_{k'} = n_k \\ \mu_{k'} + B_{k'} R \begin{pmatrix} B_k^{-1} (\theta_k - \mu_k) \\ u \end{pmatrix} & \text{if } n_{k'} > n_k \end{cases}$$

where $[\cdot]_1^m$ denotes the first m componenets, R fixed orthogonal matrix of order $\max\{n_k, n_{k'}\}$, and $u \sim q_{n_{k'} - n_k}(u)$.

Outline

Setting Up The Model

- Model is given by,

$$y = \sum_{i=1}^{n} x_i b_i + \varepsilon$$

- $n \leq m$ where m is the number of predictors, x_i's are the inputs, y is the output (what we measure), ε is the measurment noise.
- We wish to estimate $\{n, b_1, .., b_n\}$ using Bayesian modeling, given k independent measurments.
- Next, we define relevant distributions.

Setting Up The Model

- Model is given by,

$$y = \sum_{i=1}^{n} x_i b_i + \varepsilon$$

- $n \leq m$ where m is the number of predictors, x_i's are the inputs, y is the output (what we measure), ε is the measurment noise.
- We wish to estimate $\{n, b_1, .., b_n\}$ using Bayesian modeling, given k independent measurments.
- Next, we define relevant distributions.

Setting Up The Model

- Model is given by,

$$y = \sum_{i=1}^{n} x_i b_i + \varepsilon$$

- $n \leq m$ where m is the number of predictors, x_i's are the inputs, y is the output (what we measure), ε is the measurment noise.
- We wish to estimate $\{n, b_1, .., b_n\}$ using Bayesian modeling, given k independent measurments.
- Next, we define relevant distributions.

Setting Up The Model

- Model is given by,

$$y = \sum_{i=1}^{n} x_i b_i + \varepsilon$$

- $n \leq m$ where m is the number of predictors, x_i's are the inputs, y is the output (what we measure), ε is the measurment noise.
- We wish to estimate $\{n, b_1, .., b_n\}$ using Bayesian modeling, given k independent measurments.
- Next, we define relevant distributions.

Setting Up The Model (Cont'd)

- The likelihood is given by,

$$L(y|n,b) = \left(\frac{1}{\sqrt{2\pi\sigma_0^2}}\right)^k \cdot \exp\left\{-\frac{\left\|y - X^T b\right\|^2}{2\sigma_0^2}\right\}$$

- The prior over b is given by,

$$p(b|n) = \left(\frac{1}{\sqrt{2\pi\sigma_p^2}}\right)^n \cdot \exp\left\{-\frac{\left\|b-\mu\right\|^2}{2\sigma_p^2}\right\}$$

- The prior over n is given by,

$$p(n) = \frac{1}{m}$$

Setting Up The Model (Cont'd)

- The likelihood is given by,

$$L(y|n,b) = \left(\frac{1}{\sqrt{2\pi\sigma_0^2}}\right)^k \cdot \exp\left\{-\frac{\|y - X^T b\|^2}{2\sigma_0^2}\right\}$$

- The prior over b is given by,

$$p(b|n) = \left(\frac{1}{\sqrt{2\pi\sigma_p^2}}\right)^n \cdot \exp\left\{-\frac{\|b - \mu\|^2}{2\sigma_p^2}\right\}$$

- The prior over n is given by,

$$p(n) = \frac{1}{m}$$

Setting Up The Model (Cont'd)

- The likelihood is given by,

$$L(y|n,b) = \left(\frac{1}{\sqrt{2\pi\sigma_0^2}}\right)^k \cdot \exp\left\{-\frac{\|y - X^T b\|^2}{2\sigma_0^2}\right\}$$

- The prior over b is given by,

$$p(b|n) = \left(\frac{1}{\sqrt{2\pi\sigma_p^2}}\right)^n \cdot \exp\left\{-\frac{\|b - \mu\|^2}{2\sigma_p^2}\right\}$$

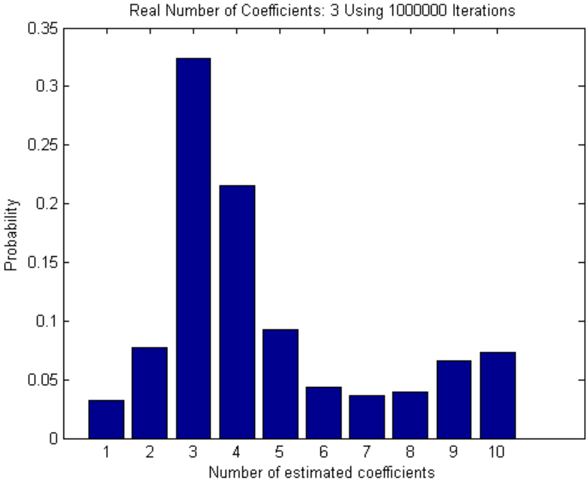
- The prior over n is given by,

$$p(n) = \frac{1}{m}$$

Outline

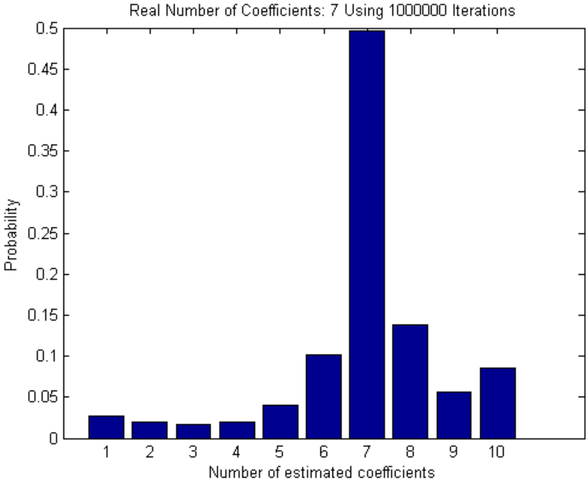
What Model Is It From?

A simulation for $n = 3$:



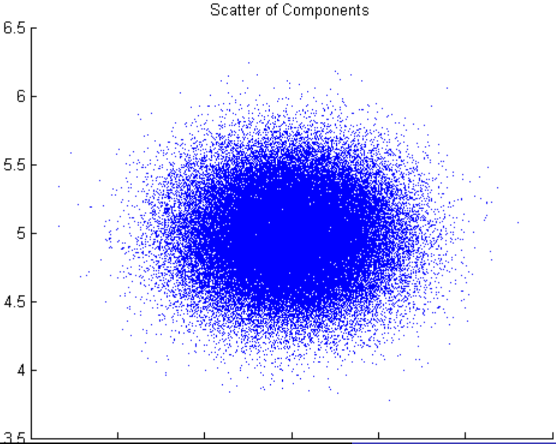
What Model Is It From?

A simulation for $n = 7$:



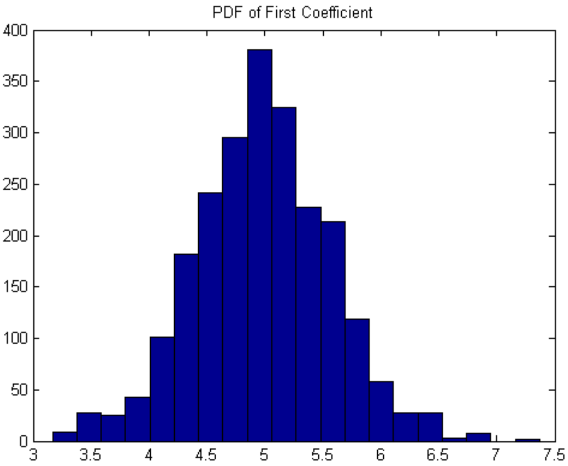
Examining a Model - PDF

A simulation for $n = 2$. Real coefficients are $\{4.99, 5.12\}$.
"Clutter" - State Space Samples.



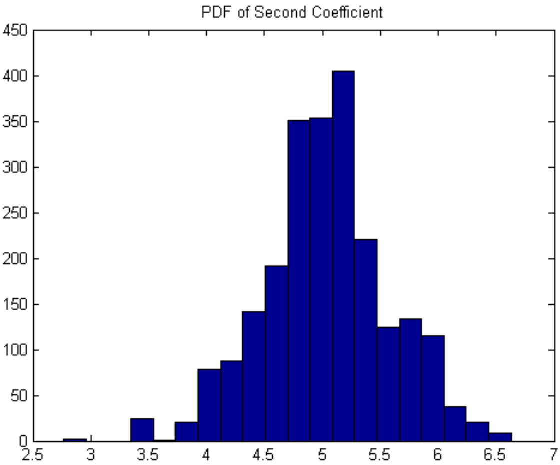
Examining a Model - PDF

A simulution for $n = 2$. Real coefficients are $\{4.99, 5.12\}$.



Examining a Model - Clutter

A simulution for $n = 2$. Real coefficients are $\{4.99, 5.12\}$.




THANK YOU