A Bayesian Approach to Diagnostics for Multivariate Control Charts

Fall Technical Conference
Philadelphia, PA
October 2017

Steven E. Rigdon
Department of Biostatistics
Saint Louis University

Robert Steward
National Geospatial-Intelligence Agency
Saint Louis, MO

Rong Pan
School of Computing, Informatics and Decision Systems Engineering
Arizona State University
Outline

• Motivating example
• MCMC Overview
• Reversible Jump MCMC Overview
• RJMCMC for Multivariate Change Point Problem
Multivariate Process Control

- Problem motivated by statistical process control (SPC)
- Multiple \( (p) \) quality characteristics are measured on each item
- Goal: Simultaneously monitor all measured quality characteristics.
- Use a multivariate control chart (e.g. Hotelling’s \( T^2 \) chart, or multivariate exponentially weighted moving average, or ...)

3
Diagnostics for Multivariate Control Chart

Model: $X_1, X_2, ..., X_\tau \sim N(\mu_\tau, \Sigma), \quad X_{\tau+1}, X_{\tau+2}, ..., X_N \sim N(\mu_{\tau+1}, \Sigma)$

If the multivariate chart signals a change (point above upper control limit on control chart), then the questions arise

1. When did the change occur?
2. Which among the $p$ components changed?
3. For those components that shifted, what are the new values for the mean?
Example (Simulated) to Illustrate the Problem

• $p = 6$

• Mean vector before the shift: $\mu_\tau = (0,0,0,0,0,0)$.

• Covariance matrix: 1’s on diagonal, 0.3’s on off-diagonal

• First 79 data points in control.

• At time 80, process mean shifts to $\mu_{\tau+1} = (0,0,0,0.75,2.00)$.

• Monitor process using Hotelling $T^2$. 
Figure: A $T^2$ control chart applied to simulated data. A change-point to the mean vector occurs at time point 80 and the control chart signals an alarm at the 99% confidence level (UCL=16.8) at time point 91.
Figure: A $T^2$ control chart applied to simulated data. A change-point to the mean vector occurs at time point 80 and the control chart signals an alarm at the 99% confidence level (UCL=16.8) at time point 91.
Now ... diagnostics. Which components shifted?

• There are $2^6 = 64$ possible models
  
  $M_1$: No change
  
  $M_2$: Component 1 mean changes
  
  $M_3$: Component 2 mean changes

  ..................

  $M_{22}$: Components 5 and 6 mean changes

  ..................

  $M_{64}$: All component means change
Now ... diagnostics. Which components shifted?

- There are $2^6 = 64$ possible models
  - $M_1$: No change
  - $M_2$: Component 1 mean changes
  - $M_3$: Component 2 mean changes
  - $M_{22}$: Components 5 and 6 mean changes **TRUE MODEL**
  - $M_{64}$: All component means change
Posterior Probability for Change Point $\tau$
Posterior Probability for Change Point $\tau$

**Histogram of Posterior Probabilities**

- **Model 22:** Components 5 and 6. Post. Prob. = 0.45
- **Model 7:** Component 6. Post. Prob. = 0.17
Posterior Probability for $\tau$
Posterior Probability for $\tau$

Posterior Mode is $\tau = 80$
Joint Posterior of Model and Change Point (with jitter)
### Estimate of Means after Change

<table>
<thead>
<tr>
<th>Component</th>
<th>Post-change mean estimate</th>
<th>95% credible interval</th>
<th>True value</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>0.71</td>
<td>(0.65,0.91)</td>
<td>0.75</td>
</tr>
<tr>
<td>6</td>
<td>1.96</td>
<td>(1.73,2.10)</td>
<td>2.00</td>
</tr>
</tbody>
</table>
Reversible Jump Markov Chain Monte Carlo (RJMC)

- Often used for model selection
- Used when parameter space for models has varying dimension
Overview of MCMC (Metropolis-Hastings)

• $X$ has pdf $f(x|\theta)$, $\theta$ has prior $p(\theta)$

• To simulate from the posterior $p(\theta|x)$

1. Start with $\theta^{(0)}$. Set $k = 1$
2. Simulate a proposal $\theta^*$ from proposal distribution $g()$
3. Accept the move to proposal with probability
   $$\alpha = \min\left(1, \frac{g(\theta^*)p(x|\theta^*)}{g(\theta^{(k-1)})p(x|\theta^{(k-1)})}\right)$$
4. $\theta^{(k)} = \theta^*$ w/prob $\alpha$, and $\theta^{(k)} = \theta^{(k-1)}$ w/prob $1 - \alpha$
5. Repeat Steps 2-4 creating a sequence $\theta^{(0)}, \theta^{(1)}, \theta^{(2)}, ...$
Overview of MCMC (Metropolis-Hastings)

• $X$ has pdf $f(x|\theta)$, $\theta$ has prior $p(\theta)$

• To simulate from the posterior $p(\theta|x)$
  1. Start with $\theta^{(0)}$. Set $k = 1$
  2. Simulate a proposal $\theta^*$ from proposal distribution $g()$
  3. Accept the move to proposal with probability
     \[
     \alpha = \min \left( 1, \frac{g(\theta^*)p(x|\theta^*)}{g(\theta^{(k-1)})p(x|\theta^{(k-1)})} \right)
     \]
  4. $\theta^{(k)} = \theta^*$ w/ prob $\alpha$, and $\theta^{(k)} = \theta^{(k-1)}$ w/ prob $1 - \alpha$
  5. Repeat Steps 2-4 creating a sequence $\theta^{(0)}, \theta^{(1)}, \theta^{(2)}, ...$

THEOREM
The steady state distribution of the sequence $\theta^{(0)}, \theta^{(1)}, \theta^{(2)}, ...$ is the posterior distribution $p(\theta|x)$. 
Example of MCMC in 1-dim Change Point Problem

\[ X_1, X_2, \ldots, X_\tau \sim N(0,1) \]
\[ X_{\tau+1}, X_{\tau+2}, \ldots, X_N \sim N(\mu, 1) \]

Simulated data with \( \tau = 40 \) and \( \mu = 1 \)

Priors: \( \tau \sim DU(1, N) \)
\( \mu \sim N(0, \text{large var.}) \)
Dimensions Vary in Multivariate Case

• Suppose \( p = 2 \). IC mean \( \mu_0 = [0, 0]' \). IC covariance \( \Sigma = I \).

• Possible Models
  1. No change. Parameters: none
  2. Only component 1 shifts. New mean is \( \mu_1 = [\mu_{21}, 0]' \)
     Parameters: \( \tau_2, \mu_{21} \)
  3. Only component 2 shifts. New mean is \( \mu_1 = [0, \mu_{32}]' \)
     Parameters: \( \tau_3, \mu_{32} \)
  4. Both components shift. New mean is \( \mu_1 = [\mu_{41}, \mu_{42}]' \)
     Parameters: \( \tau_4, \mu_{41}, \mu_{42} \)
Dimensions Vary in Multivariate Case

• Suppose $p = 2$. IC mean $\mu_0 = [0,0]'$. IC covariance $\Sigma = I$.

• Possible Models

  1. No change. Parameters: none

  2. Only component 1 shifts. New mean is $\mu_1 = [\mu_{21}, 0]'$
     Parameters: $\tau_2, \mu_{21}$

  3. Only component 2 shifts. New mean is $\mu_1 = [0, \mu_{32}]'$
     Parameters: $\tau_3, \mu_{32}$

  4. Both components shift. New mean is $\mu_1 = [\mu_{41}, \mu_{42}]'$
     Parameters: $\tau_4, \mu_{41}, \mu_{42}$
Dimensions of Parameter Space Vary

• The number of unknown parameters varies, depending on the model.

• “The number of things you don’t know is one of the things you don’t know.” (Hastie, 1995)

Overview of Reversible Jump MCMC

• Consider models $M_k, k = 1, 2, \ldots, L$.

• Model $M_k$ has parameter $\theta_k$.

• The model specific posterior distribution is

$$
\pi_k(\theta_k | D, M_k) = \frac{p_0(\theta_k | M_k) L(D | \theta_k, M_k)}{p_k(D | M_k)}
$$
Treat the model as an additional parameter.

• Treat the model $M_k$ as an additional parameter.

• $S_k$ denotes parameter space for $M_k$

• $S = \bigcup_{k=1}^{L} \{M_k\} \times S_k$

• Goal: Sample from $S$ in an MCMC fashion to produce a chain that converges to the posterior distribution

$$
\pi(M_k, \theta_k | D) \propto p_0(M_k)p_0(\theta_k | M_k)L(D | M_k, \theta_k)
$$
Overview of RJMCMC

• Propose move from model $M_k$ with parameter $x$ to model $M_{k'}$ with parameter $x'$

• Chain must be aperiodic and irreducible, and the detailed balance equation must be satisfied:

$$\pi(x)j(M_k|M_{k'})g(u)\alpha(x, x') = \pi(x')j(M_{k'}|M_k)g'(u')\alpha(x', x) \left| \frac{\partial (x', u')}{\partial (x, u)} \right|$$

$u$ is a “padding” variable that accounts for the difference in dimension between proposed models.
Acceptance Probability Is

\[ \alpha(x, x') = \min \left( 1, \frac{\pi(x') j(M_{k'}, M_k) g'(u')}{\pi(x) j(M_k | M_{k'}) g(u)} \right) \]

Dimension Matching

\[ n_k = \text{dimension of } x \]
\[ r_k = \text{dimension of } u \]
\[ n'_k = \text{dimension of } x' \]
\[ r'_k = \text{dimension of } u' \]

\[ n_k + r_k = n'_k + r'_k \]
RJMCMC Algorithm (Sketch)

1. Choose initial conditions (state) \( x_0 = (M_{k_0}, \mu_{k_0}, \tau_{k_0}) \)

2. Within-model MH update of \((\mu_{k_0}, \tau_{k_0})\)

3. Propose jump to model \( M_{k'} \) with PMF \( j(M_{k'}, \middle| M_k) \)

4. If jumping to a model with more parameters, simulate \( u \)

5. Accept move to \( x'_0 = (M_{k_0}, \mu_{k_0}, \tau_{k_0}) \) with probability \( \alpha(x_0, x'_0) \)

6. Repeat steps 2 – 5 until MC convergence. Then run additional simulations to explore posterior.
Possible implementation of RJMCMC

• Within-model MH (standard stuff)

• $\tau \sim \text{DU}(\text{centered at current } \tau)$

• Model DU on all models that add one component or remove one component.

  Example: $p = 4, M = \{1,4\}$:

  Possible proposed models: $\{1,2,4\}, \{1,3,4\}, \{1\}, \{4\}$
Smaller model to larger model ...

- Suppose we are in state $M = \{1\}$ and we propose to move to model $\{1,4\}$.

Current state (within model 2, $\{1\}$)  \[
\tilde{\theta}_k = [\mu_{2,\tau+1}, \tau]
\]

maps to (within model 8, $\{1,4\}$)  \[
\tilde{\theta}_{k'} = [\mu_{8,\tau+1}, \mu_{8,\tau+1}, \tau]
\]
Suppose we are in state $M = \{1\}$ and we propose to move to model $\{1,4\}$.

Current state (within model 2, $\{1\}$) \[ \tilde{\theta}_k = [\mu_{2,\tau+1}, \tau] \]

maps to (within model 8, $\{1,4\}$) \[ \tilde{\theta}_{k'} = [\mu_{8,\tau+1}, \mu_{8,\tau+1}, \mu_{4,\tau+1}, \tau] \]
Smaller model to larger model ...

- Suppose we are in state $M = \{1\}$ and we propose to move to model $\{1, 4\}$.

$$u \sim N (\text{last visit}, \sigma_u^2)$$

Current state (within model 2, $\{1\}$)

$$\tilde{\theta}_k = [\mu_{2, \tau+1}^1, u, \tau]$$

maps to (within model 8, $\{1, 4\}$)

$$\tilde{\theta}_{k'} = [\mu_{8, \tau+1}^1, \mu_{8, \tau+1}^4, \tau]$$
Larger model to smaller model ...

• Suppose we are in state $M = \{1, 4\}$ and we propose to move to model $\{1\}$.

Current state (within model 8, $\{1, 4\}$)  
\[
\tilde{\theta}_k = \begin{bmatrix}
\mu_{8,\tau+1},^{1} \mu_{8,\tau+1},^{4} \tau
\end{bmatrix}
\]

maps to (within model 2, $\{1\}$)  
\[
\tilde{\theta}_{kr} = \begin{bmatrix}
\mu_{2,\tau+1},^{1} u, \tau
\end{bmatrix}
\]
Model

Possible Assumptions

\[ X_1, X_2, \ldots, X_\tau \sim N(\mu_\tau, \Sigma) \]
\[ X_{\tau+1}, X_{\tau+2}, \ldots, X_N \sim N(\mu_{\tau+1}, \Sigma) \]

1. known: \( \mu_\tau, \Sigma \)
   unknown: \( \tau, \mu_{\tau+1} \)
   (unrealistic, but easy to explain)

2. known: \( \Sigma \)
   unknown: \( \tau, \mu_\tau, \mu_{\tau+1} \)
   (stepping stone)

3. known: nothing
   unknown: \( \tau, \mu_\tau, \mu_{\tau+1}, \Sigma \)
   (realistic, but messy)
Summary

• Motivated by multivariate SPC

• Single model to address
  
  When did the change occur?
  Which components changed?
  What are the new means?

• Extensions to multiple change points by (even messier) MCMC

Binary Segmentation
But ... SPC is often looking only for single change