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 Pont 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 …)

Diagnostics for Multivariate Control Chart

Model: $\boldsymbol{X}_1, \boldsymbol{X}_2, \dots, \boldsymbol{X}_\tau \sim N(\boldsymbol{\mu}_\tau, \Sigma), \quad \boldsymbol{X}_{\tau+1}, \boldsymbol{X}_{\tau+2}, \dots, \boldsymbol{X}_N \sim N(\boldsymbol{\mu}_{\tau+1}, \Sigma)$

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 $\mu_{\tau+1} = (0,0,0,0,0.75,2.00)$.
- Monitor process using Hotelling T^2.

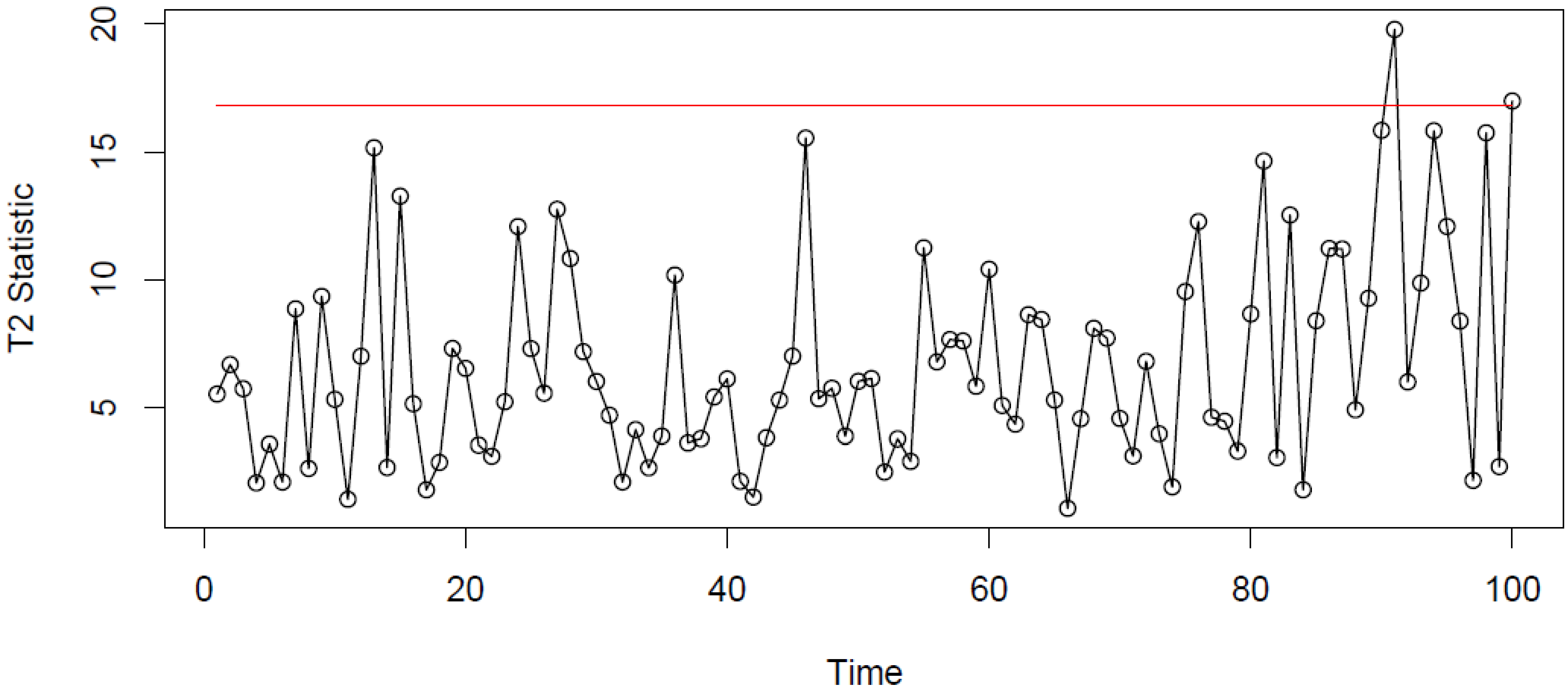


Figure : A T^2 control chart applied to simulated data. A change-point to the mean vector occurs at time point 80 and the control chart signals an alarm at the 99% confidence level (UCL=16.8) at time point 91.

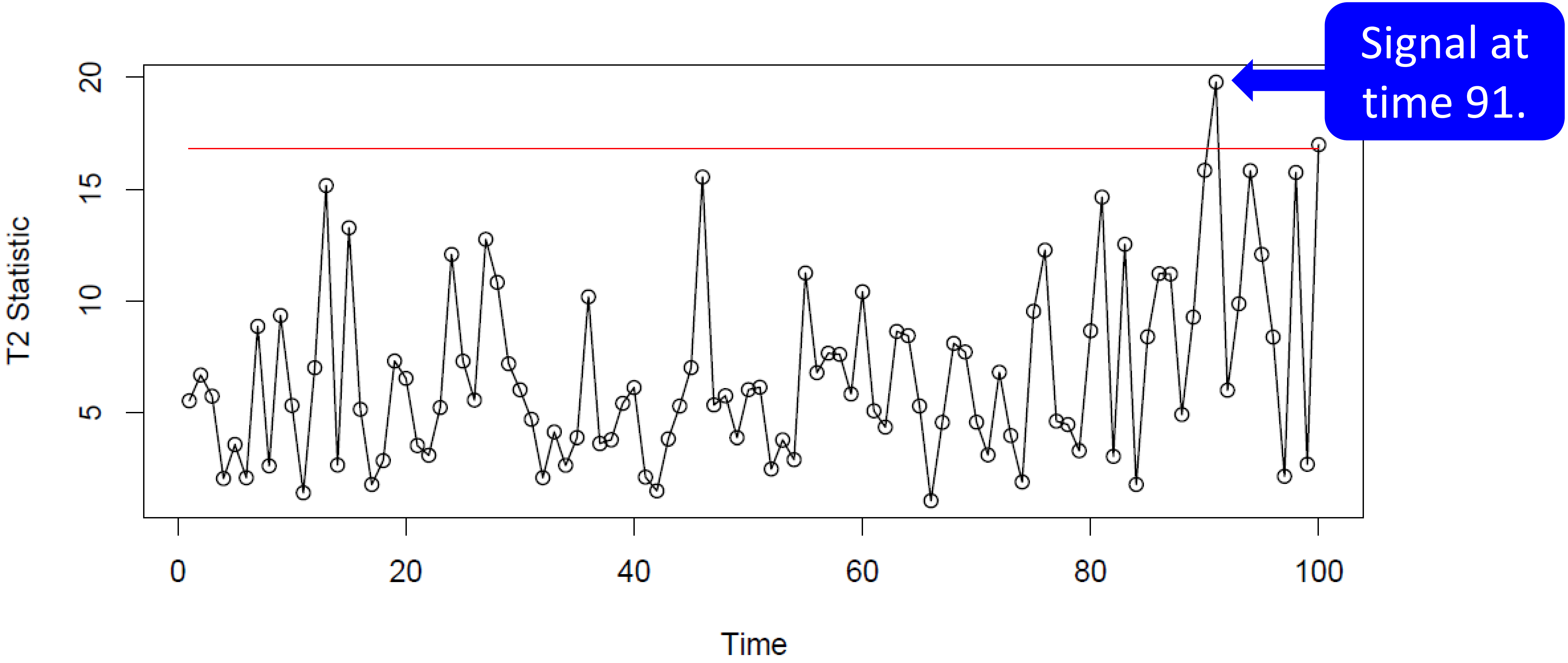


Figure : A T^2 control chart applied to simulated data. A change-point to the mean vector occurs at time point 80 and the control chart signals an alarm at the 99% confidence level (UCL=16.8) at time point 91.

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

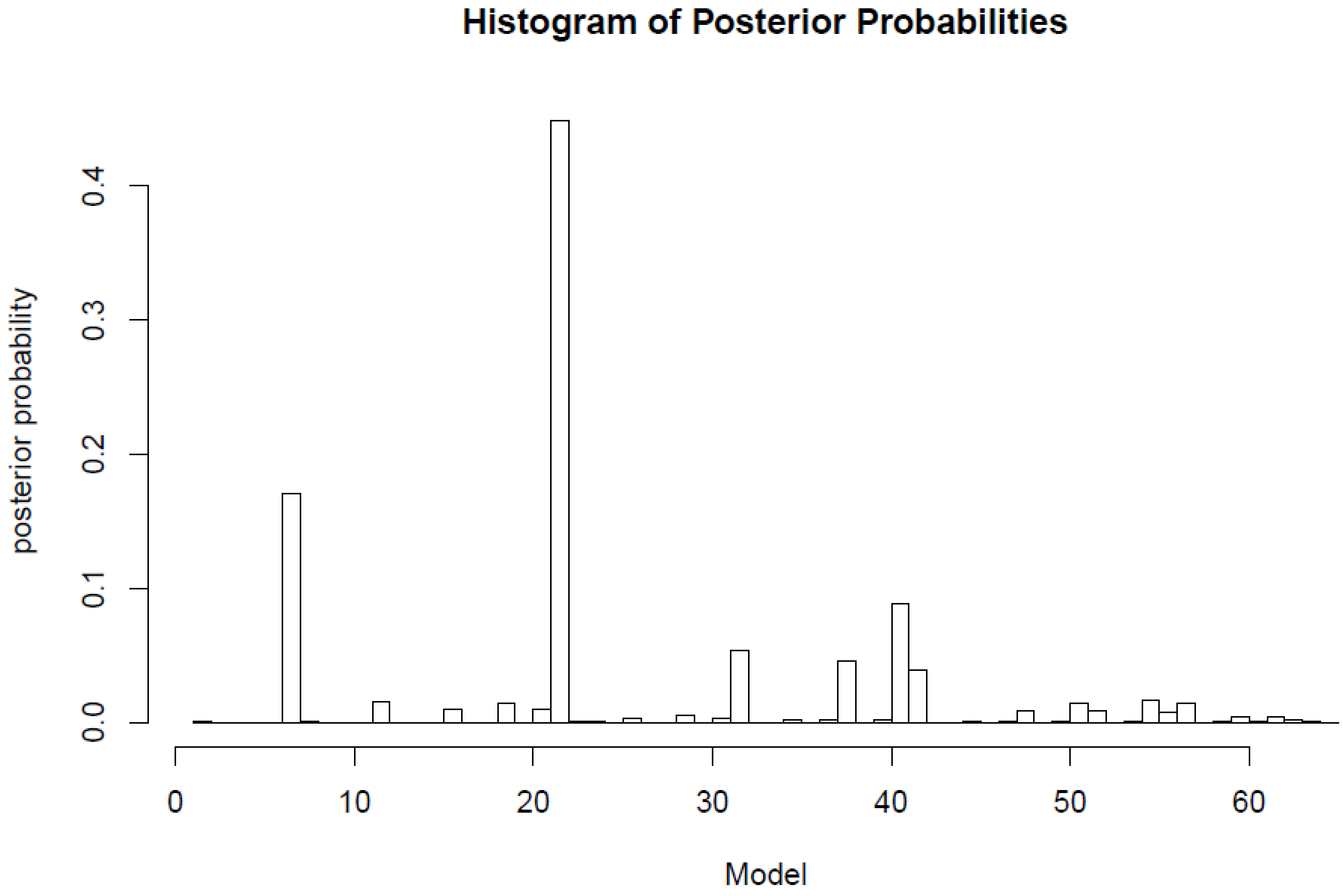
$\cdots\cdots\cdots\cdots\cdots$

M_{22}: Components 5 and 6 mean changes ←TRUE MODEL

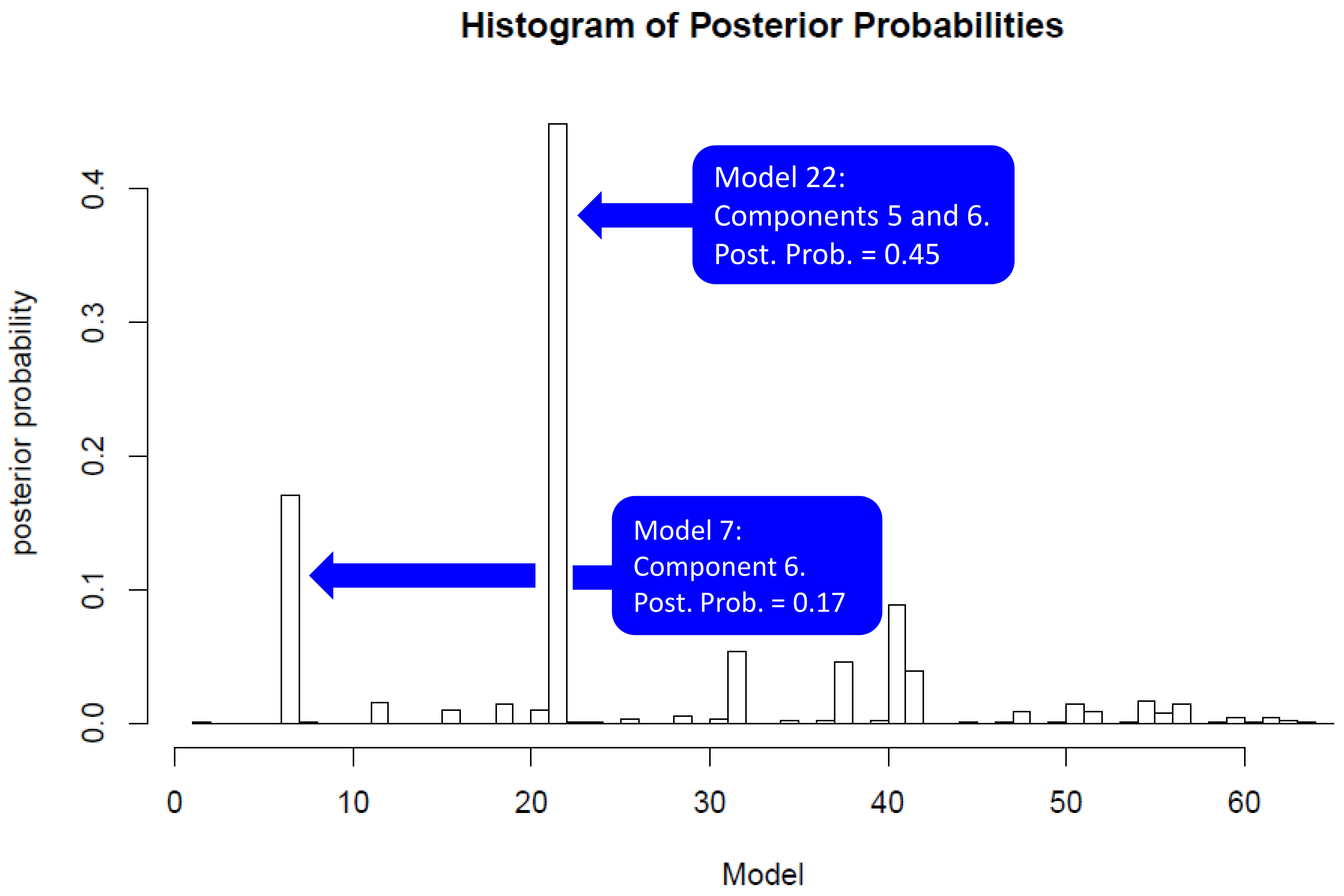
$\cdots\cdots\cdots\cdots\cdots$

M_{64}: All component means change

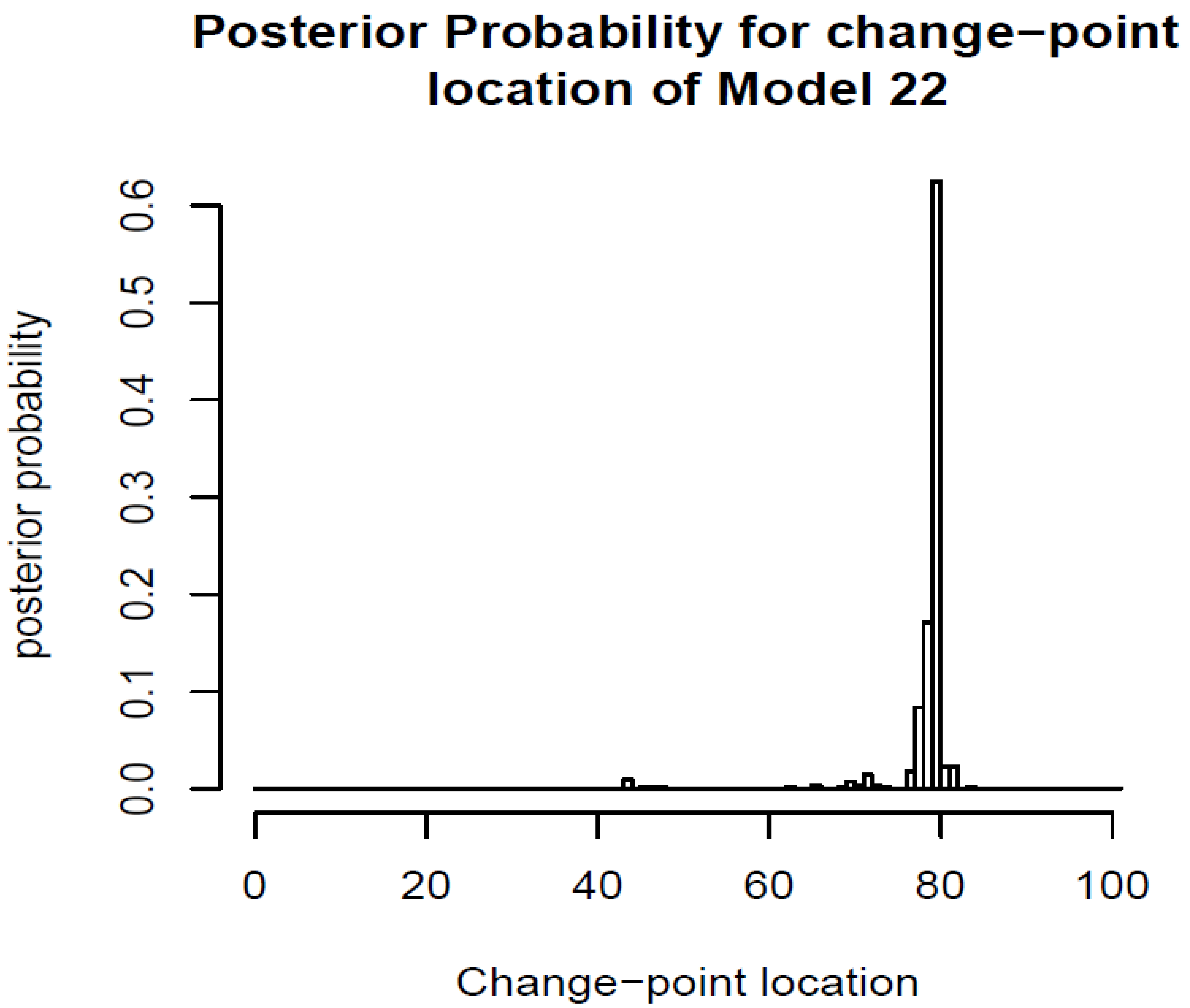
Posterior Probability for Change Point τ



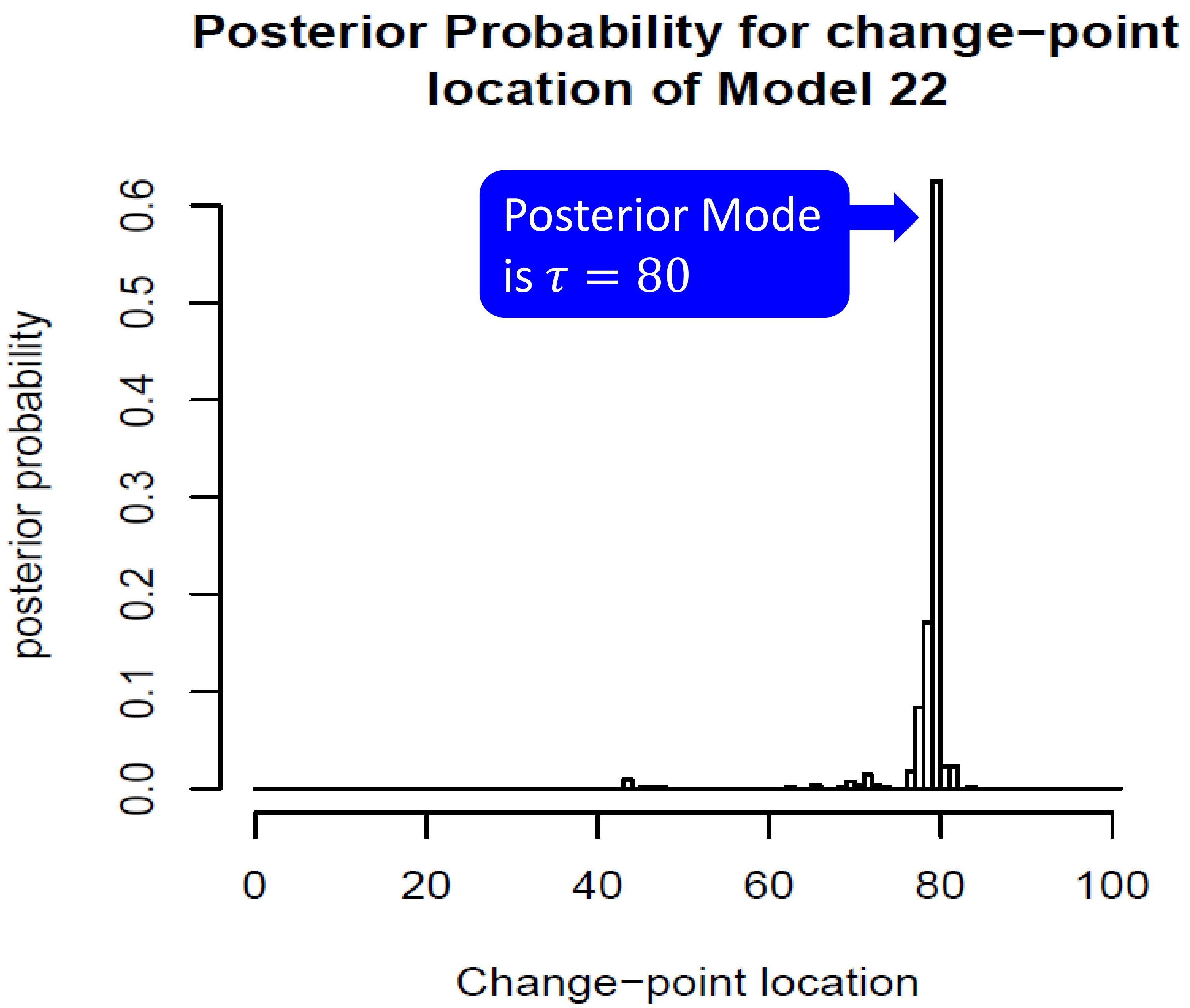
Posterior Probability for Change Point τ



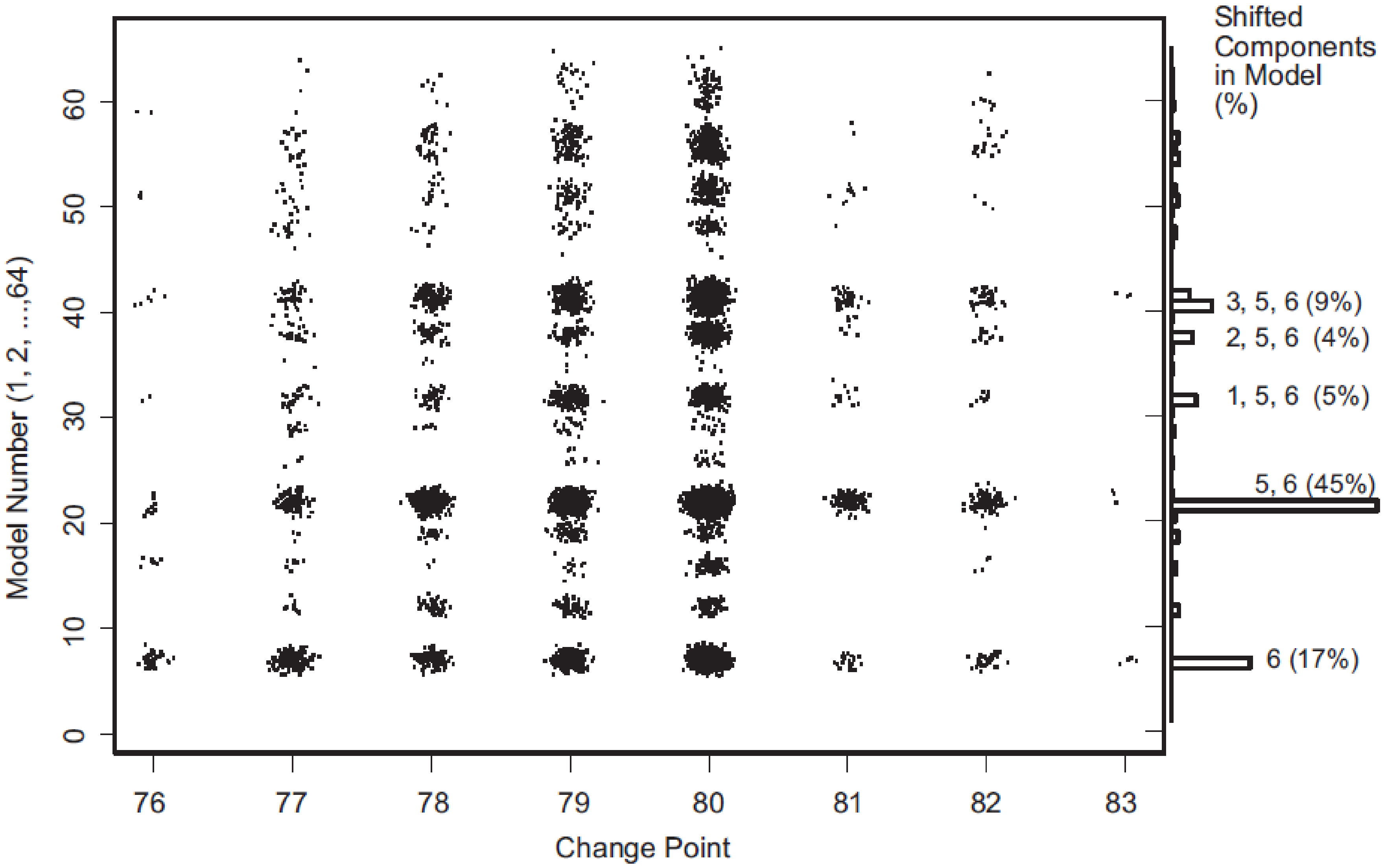
Posterior Probability for τ



Posterior Probability for τ



Joint Posterior of Model and Change Point (with jitter)



Estimate of Means after Change

Component	Post-change mean estimate	95% credible interval	True value
5	0.71	(0.65,0.91)	0.75
6	1.96	(1.73,2.10)	2.00

Reversible Jump Markov Chain Monte Carlo (RJMCMC)

- Often used for model selection
- Used when parameter space for models has varying dimension

Overview of MCMC (Metropolis-Hastings)

- $\boldsymbol{X}$ has pdf $f(\boldsymbol{x}|\boldsymbol{\theta})$, $\boldsymbol{\theta}$ has prior $p(\boldsymbol{\theta})$
- To simulate from the posterior $p(\boldsymbol{\theta}|\boldsymbol{x})$

1. Start with $\boldsymbol{\theta}^{(0)}$. Set $k = 1$
2. Simulate a proposal $\boldsymbol{\theta}^*$ from proposal distribution $g()$
3. Accept the move to proposal with probability

$$\alpha = \min\left(1, \frac{g(\boldsymbol{\theta}^*)p(\boldsymbol{x}|\boldsymbol{\theta}^*)}{g\left(\boldsymbol{\theta}^{(k-1)}\right)p\left(\boldsymbol{x}|\boldsymbol{\theta}^{(k-1)}\right)}\right)$$

4. $\boldsymbol{\theta}^{(k)} = \boldsymbol{\theta}^*$ w/prob α, and $\boldsymbol{\theta}^{(k)} = \boldsymbol{\theta}^{(k-1)}$ w/prob $1 - \alpha$
5. Repeat Steps 2-4 creating a sequence $\boldsymbol{\theta}^{(0)}, \boldsymbol{\theta}^{(1)}, \boldsymbol{\theta}^{(2)}, \ldots$

Overview of MCMC (Metropolis-Hastings)

- $\boldsymbol{X}$ has pdf $f(\boldsymbol{x}|\boldsymbol{\theta})$, $\boldsymbol{\theta}$ has prior $p(\boldsymbol{\theta})$
- To simulate from the posterior $p(\boldsymbol{\theta}|\boldsymbol{x})$

1. Start with $\boldsymbol{\theta}^{(0)}$. Set $k = 1$
2. Simulate a proposal $\boldsymbol{\theta}^*$ from proposal distribution $g()$
3. Accept the move to proposal with probability

$$\alpha = \min\left(1, \frac{g(\boldsymbol{\theta}^*)p(\boldsymbol{x}|\boldsymbol{\theta}^*)}{g(\boldsymbol{\theta}^{(k-1)})p(\boldsymbol{x}|\boldsymbol{\theta}^{(k-1)})}\right)$$

4. $\boldsymbol{\theta}^{(k)} = \boldsymbol{\theta}^*$ w/prob α, and $\boldsymbol{\theta}^{(k)} = \boldsymbol{\theta}^{(k-1)}$ w/prob $1 - \alpha$
5. Repeat Steps 2-4 creating a sequence $\boldsymbol{\theta}^{(0)}, \boldsymbol{\theta}^{(1)}, \boldsymbol{\theta}^{(2)}, \ldots$

<u>THEOREM</u>
The steady state distribution of the sequence $\boldsymbol{\theta}^{(0)}, \boldsymbol{\theta}^{(1)}, \boldsymbol{\theta}^{(2)}, \ldots$ is the posterior distribution $p(\boldsymbol{\theta}|\boldsymbol{x})$.

Example of MCMC in 1-dim Change Point Problem

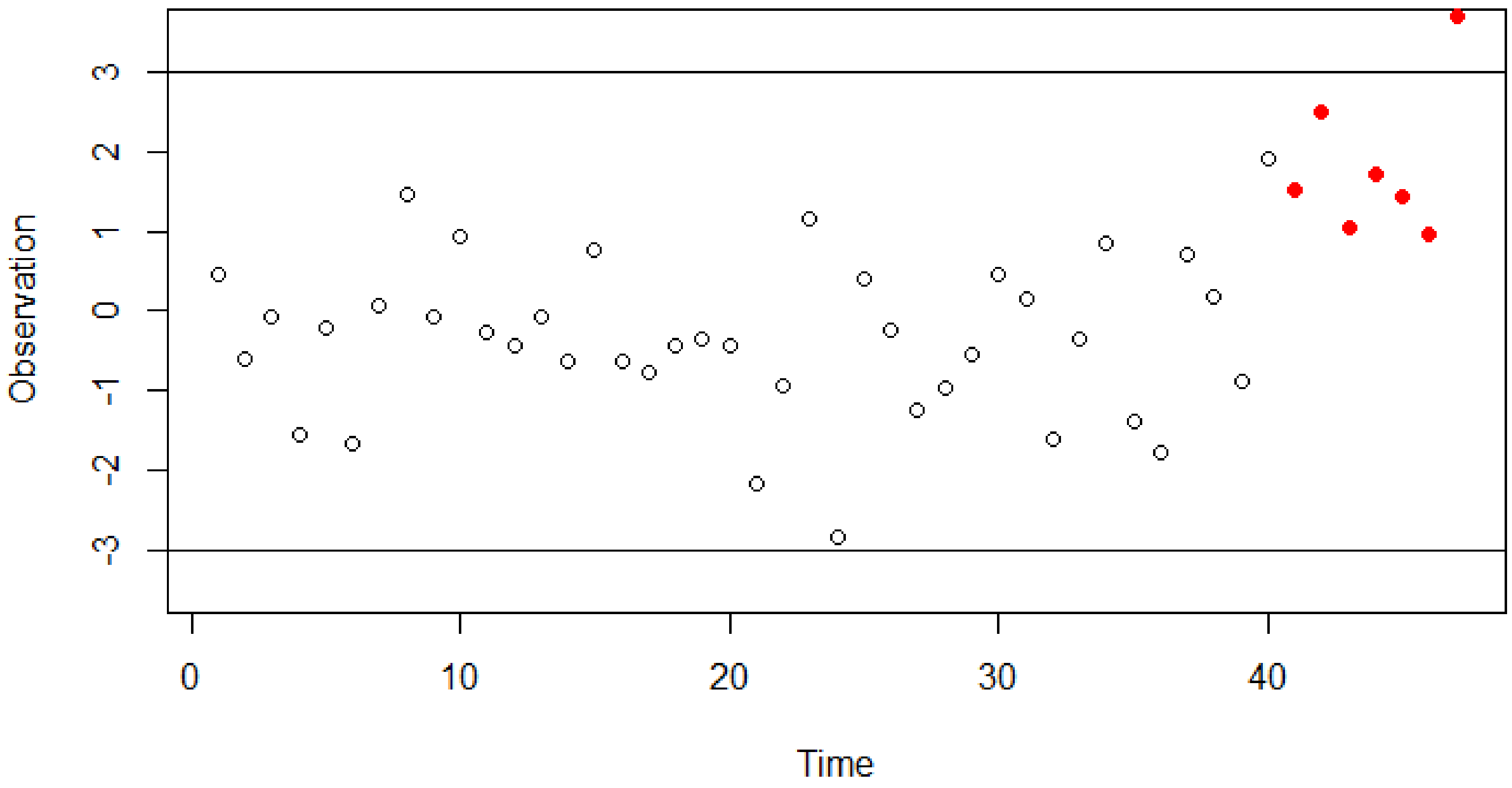
$X_1, X_2, \dots, X_\tau \sim N(0,1)$

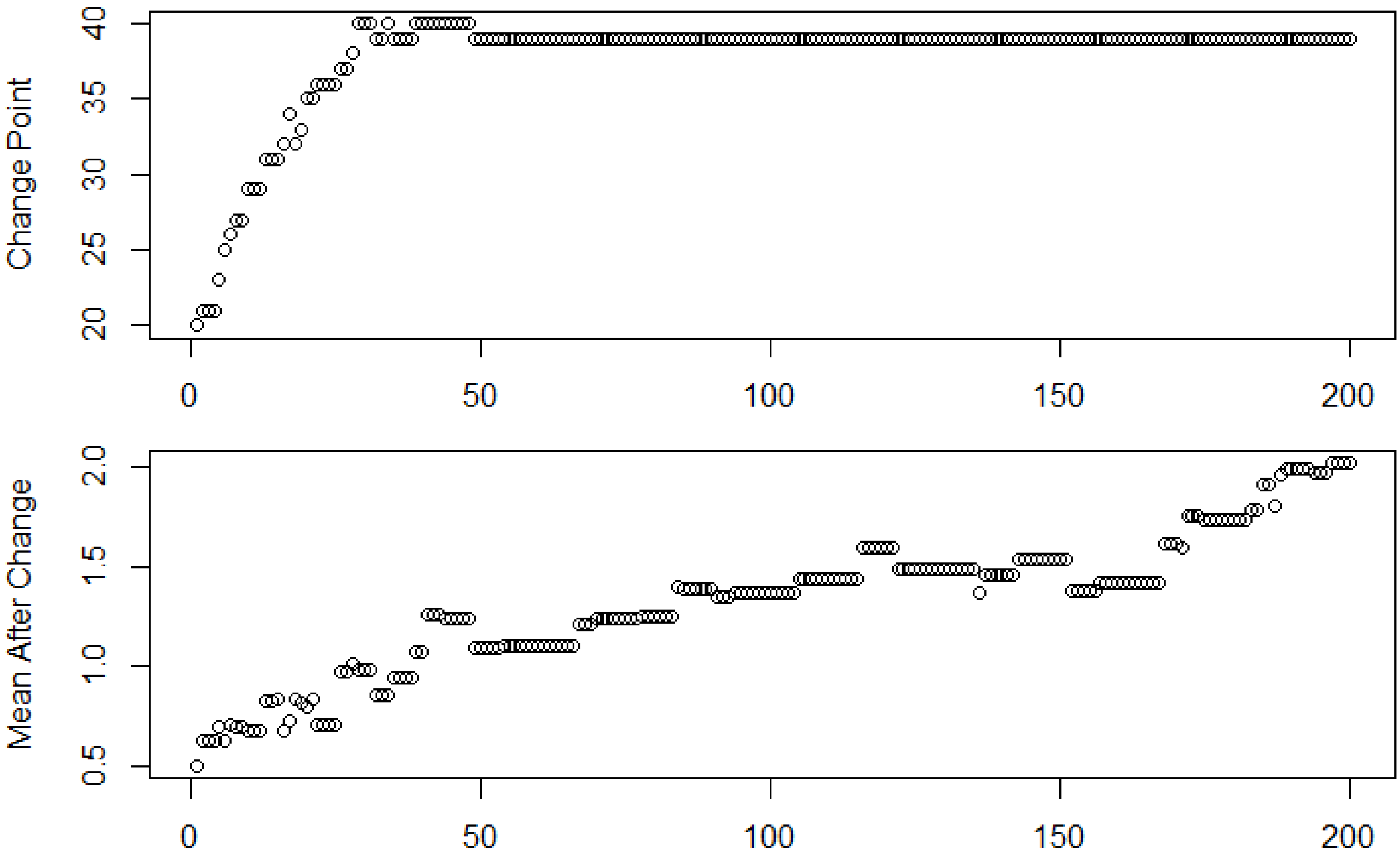
$X_{\tau+1}, X_{\tau+2}, \dots, X_N \sim N(\mu, 1)$

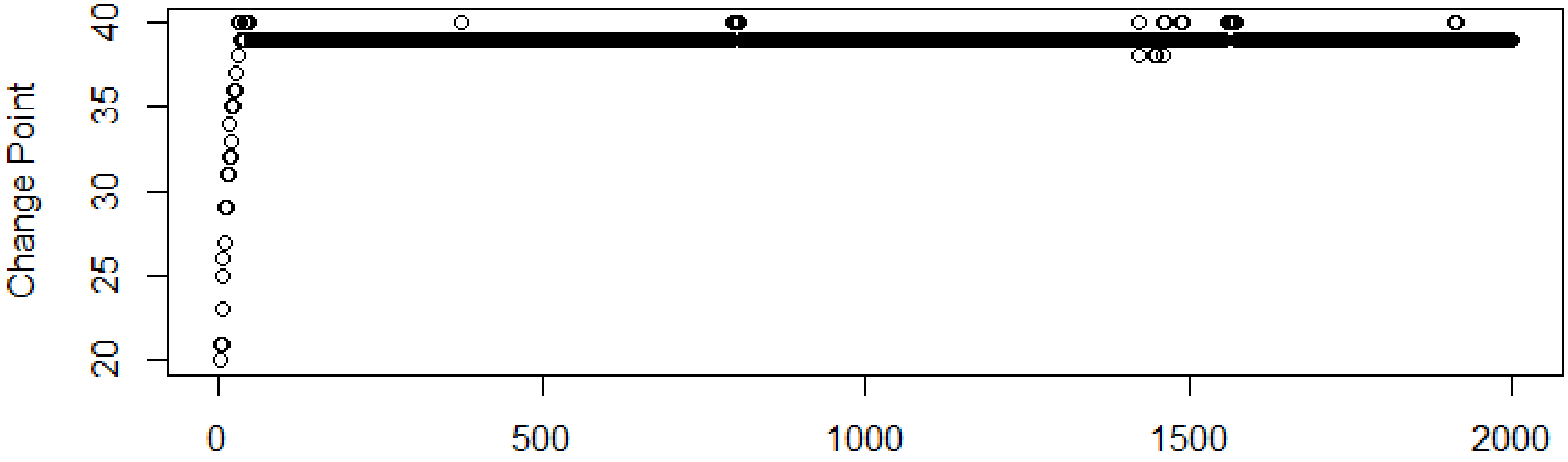
Simulated data with $\tau = 40$ and $\mu = 1$

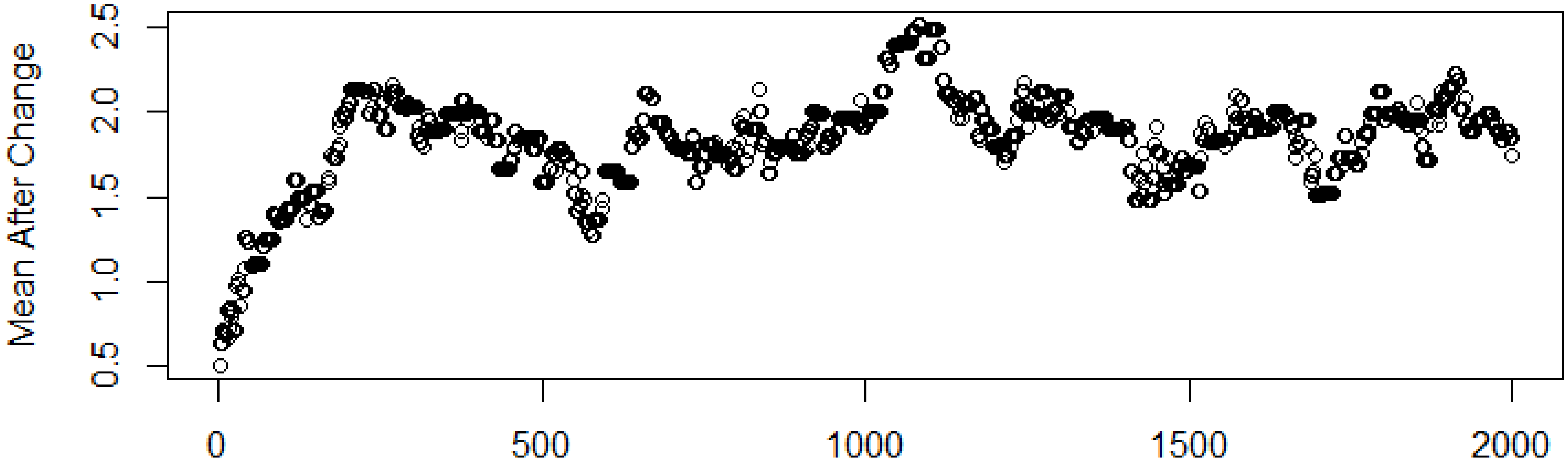
Priors: $\tau \sim DU(1, N)$

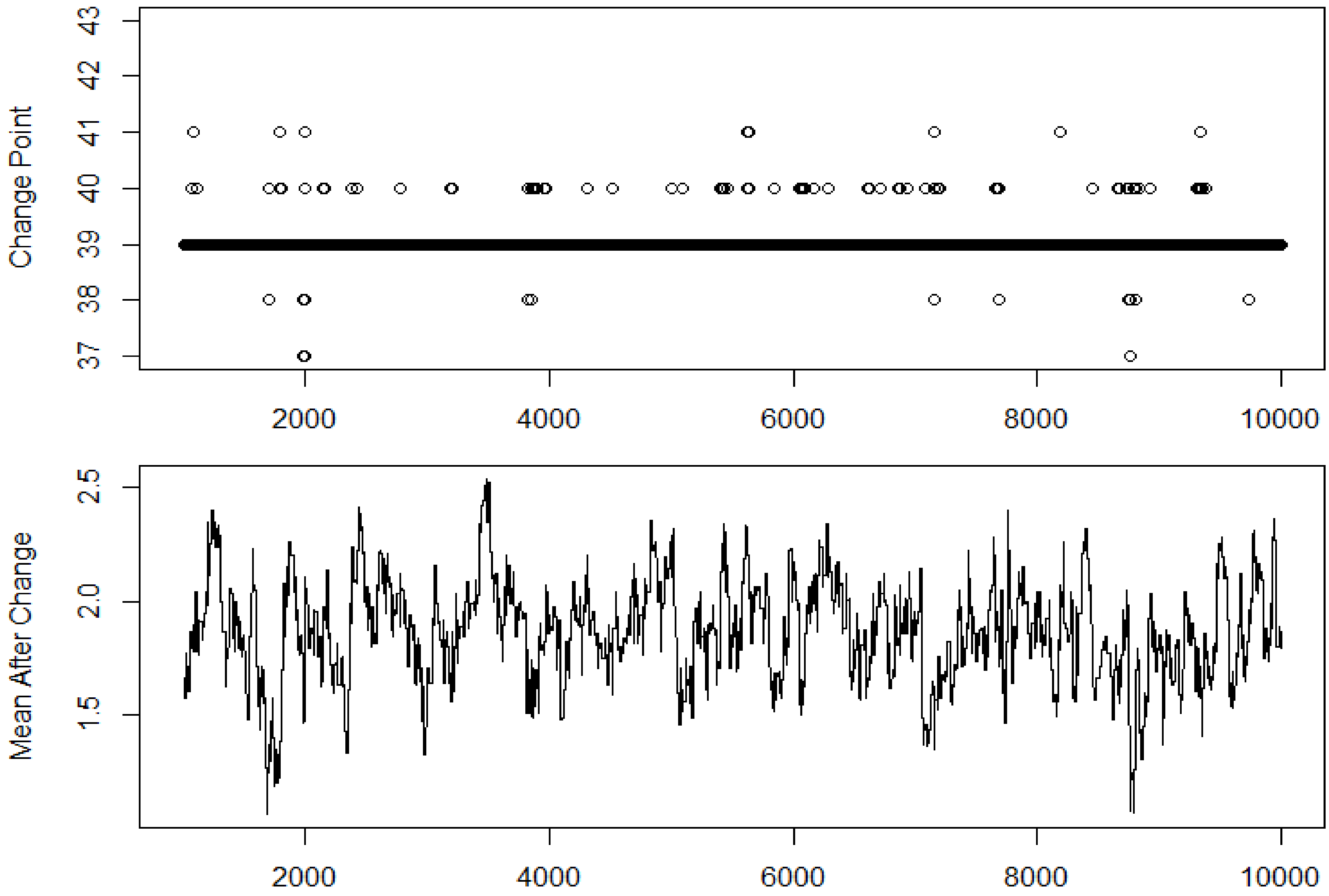
$\mu \sim N(0, \text{large var.})$

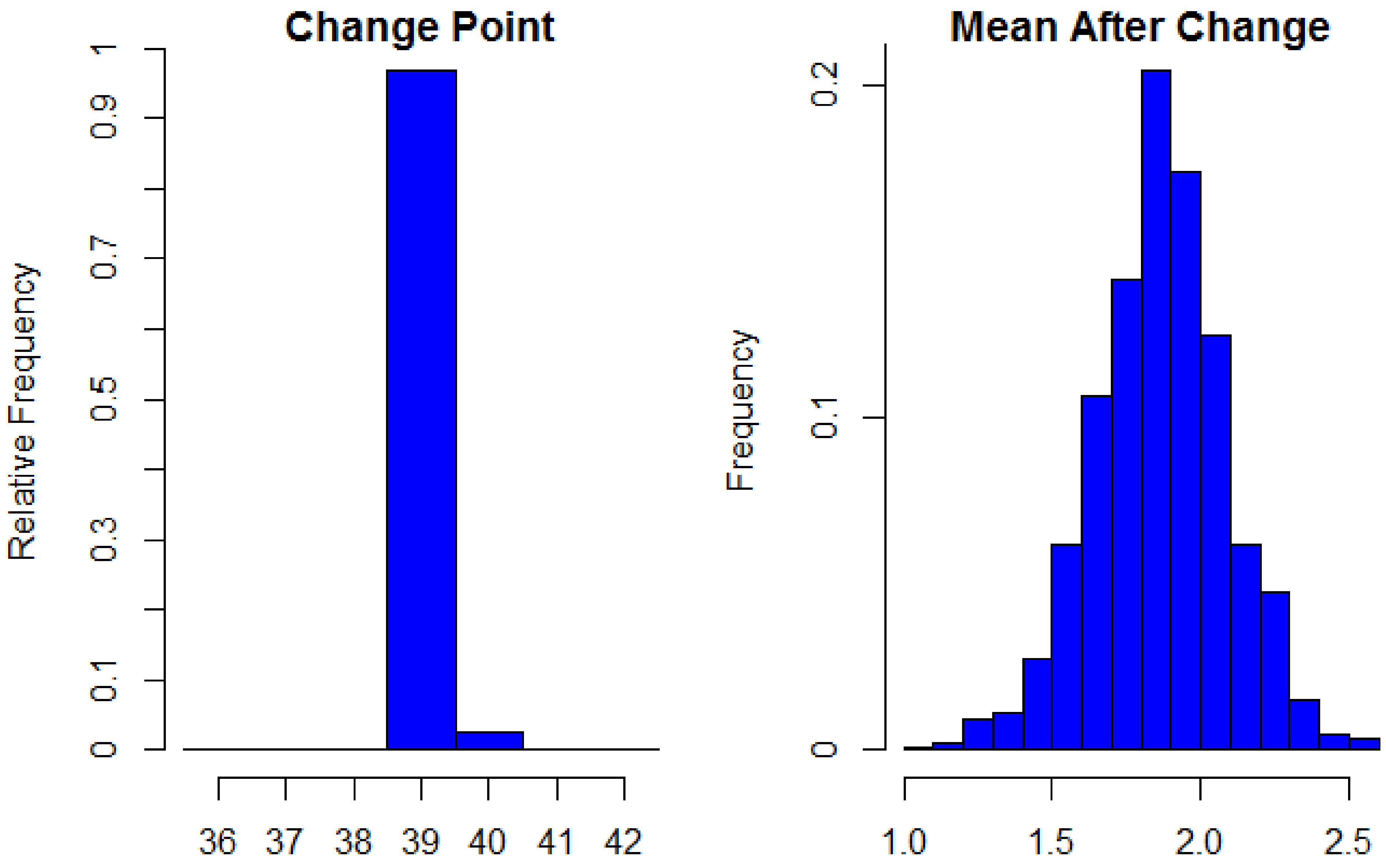



Change Point
20
25
30
35
40
0
50
100
150
200
Mean After Change
0.5
1.0
1.5
2.0
0
50
100
150
200


Change Point
20
25
30
35
40
0
500
1000
1500
2000


Mean After Change
0.5
1.0
1.5
2.0
2.5
0
500
1000
1500
2000


43
42
41
40
39
38
37
Change Point
2000
4000
6000
8000
10000
2.5
2.0
1.5
Mean After Change
2000
4000
6000
8000
10000


Change Point
Relative Frequency
0
0.1
0.3
0.5
0.7
0.9
1
36
37
38
39
40
41
42
Mean After Change
Frequency
0
0.1
0.2
1.0
1.5
2.0
2.5

Dimensions Vary in Multivariate Case

- Suppose $p = 2$. IC mean $\boldsymbol{\mu}_0 = [0,0]'$. IC covariance $\Sigma = I$.
- Possible Models
    1. No change. Parameters: none
    2. Only component 1 shifts. New mean is $\boldsymbol{\mu}_1 = [\mu_{21}, 0]'$
       Parameters: τ_2, μ_{21}
    3. Only component 2 shifts. New mean is $\boldsymbol{\mu}_1 = [0, \mu_{32}]'$
       Parameters: τ_3, μ_{32}
    4. Both components shift. New mean is $\boldsymbol{\mu}_1 = [\mu_{41}, \mu_{42}]'$
       Parameters: $\tau_4, \mu_{41}, \mu_{42}$

Dimensions Vary in Multivariate Case

- Suppose $p = 2$. IC mean $\boldsymbol{\mu}_0 = [0,0]'$. IC covariance $\Sigma = I$.
- Possible Models

1. No change. Parameters: none

   dim = 0

2. Only component 1 shifts. New mean is $\boldsymbol{\mu}_1 = [\mu_{21}, 0]'$
   Parameters: τ_2, μ_{21}

   dim = 2

3. Only component 2 shifts. New mean is $\boldsymbol{\mu}_1 = [0, \mu_{32}]'$
   Parameters: τ_3, μ_{32}

   dim = 2

4. Both components shift. New mean is $\boldsymbol{\mu}_1 = [\mu_{41}, \mu_{42}]'$
   Parameters: $\tau_4, \mu_{41}, \mu_{42}$

   dim = 3

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)
- Green (1995) suggested method for ***transdimensional*** parameter space models, and called it Reversible Jump Markov chain Monte Carlo (RJMCMC).

Overview of Reversible Jump MCMC

- Consider models $M_k, k = 1,2, \dots, L$.
- Model M_k has parameter $\boldsymbol{\theta}_k$.
- The model specific posterior distribution is

$$\pi_k(\boldsymbol{\theta}_k \,|\boldsymbol{D}, M_k\,) = \frac{p_0(\boldsymbol{\theta}_k|M_k)\, L(\boldsymbol{D}|\boldsymbol{\theta}_k, M_k)}{p_k(\boldsymbol{D}|M_k)}$$

Treat the model as an additional parameter.

- Treat the model M_k as an additional parameter.
- S_k denotes parameter space for M_k
- $S = \cup_{k=1}^{L} \{M_k\} \times S_k$
- Goal: Sample from S in an MCMC fashion to produce a chain that converges to the posterior distribution

$$\pi(M_k, \boldsymbol{\theta}_k | \boldsymbol{D}) \propto p_0(M_k) p_0(\boldsymbol{\theta}_k | M_k) L(\boldsymbol{D} | M_k, \boldsymbol{\theta}_k)$$

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

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

Dimension Matching

n_k = dimension of x

n'_k = dimension of x'

r_k = dimension of u

r'_k = dimension of u'

$$n_k + r_k = n'_k + r'_k$$

RJMCMC Algorithm (Sketch)

1. Choose initial conditions (state) $x_0 = \left(M_{k_0}, \boldsymbol{\mu}_{k_0}, \tau_{k_0}\right)$
2. Within-model MH update of $\left(\boldsymbol{\mu}_{k_0}, \tau_{k_0}\right)$
3. Propose jump to model $M_{k'}$, with PMF $j(M_{k'}|M_k)$
4. If jumping to a model with more parameters, simulate u
5. Accept move to $x_0' = \left(M_{k_0}, \boldsymbol{\mu}_{k_0}, \tau_{k_0}\right)$ 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)
- $\tau \sim \mathrm{DU}$(centered at current τ)
- 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, $\{1\}$) $\tilde{\theta}_k = [\mu^1_{2,\tau+1}, \tau]$

$$\Downarrow$$

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

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, $\{1\}$) $\tilde{\theta}_k = [\mu^1_{2,\tau+1}, \tau]$ 2-dim

$\Downarrow$

maps to (within model 8, {1,4}) $\tilde{\theta}_{k'} = [\mu^1_{8,\tau+1}, \mu^4_{8,\tau+1}, \tau]$ 3-dim

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}) $\tilde{\theta}_k = [\mu^1_{2,\tau+1}, u, \tau]$ 3-dim

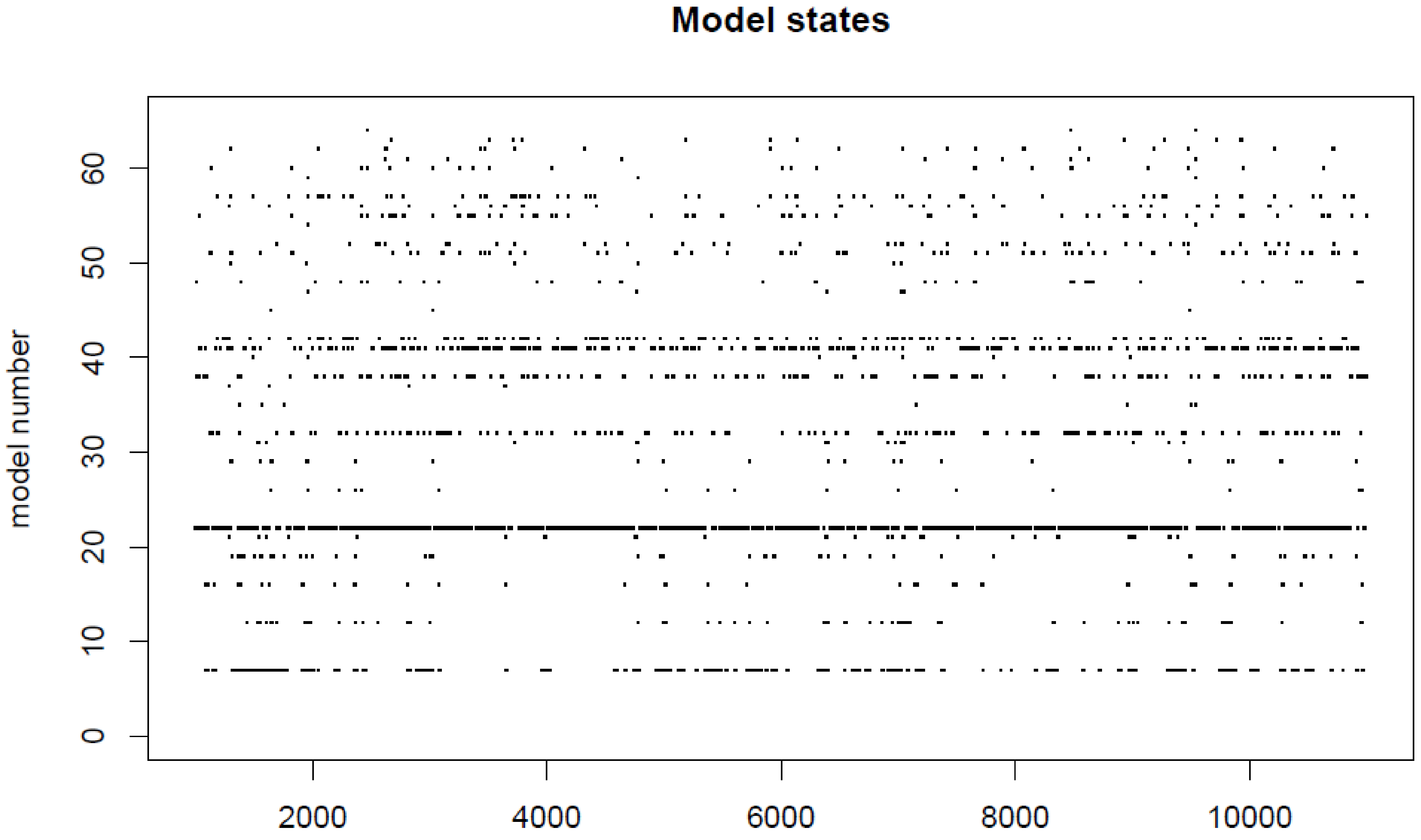
maps to (within model 8, {1,4}) $\tilde{\theta}_{k'} = [\mu^1_{8,\tau+1}, \mu^4_{8,\tau+1}, \tau]$ 3-dim

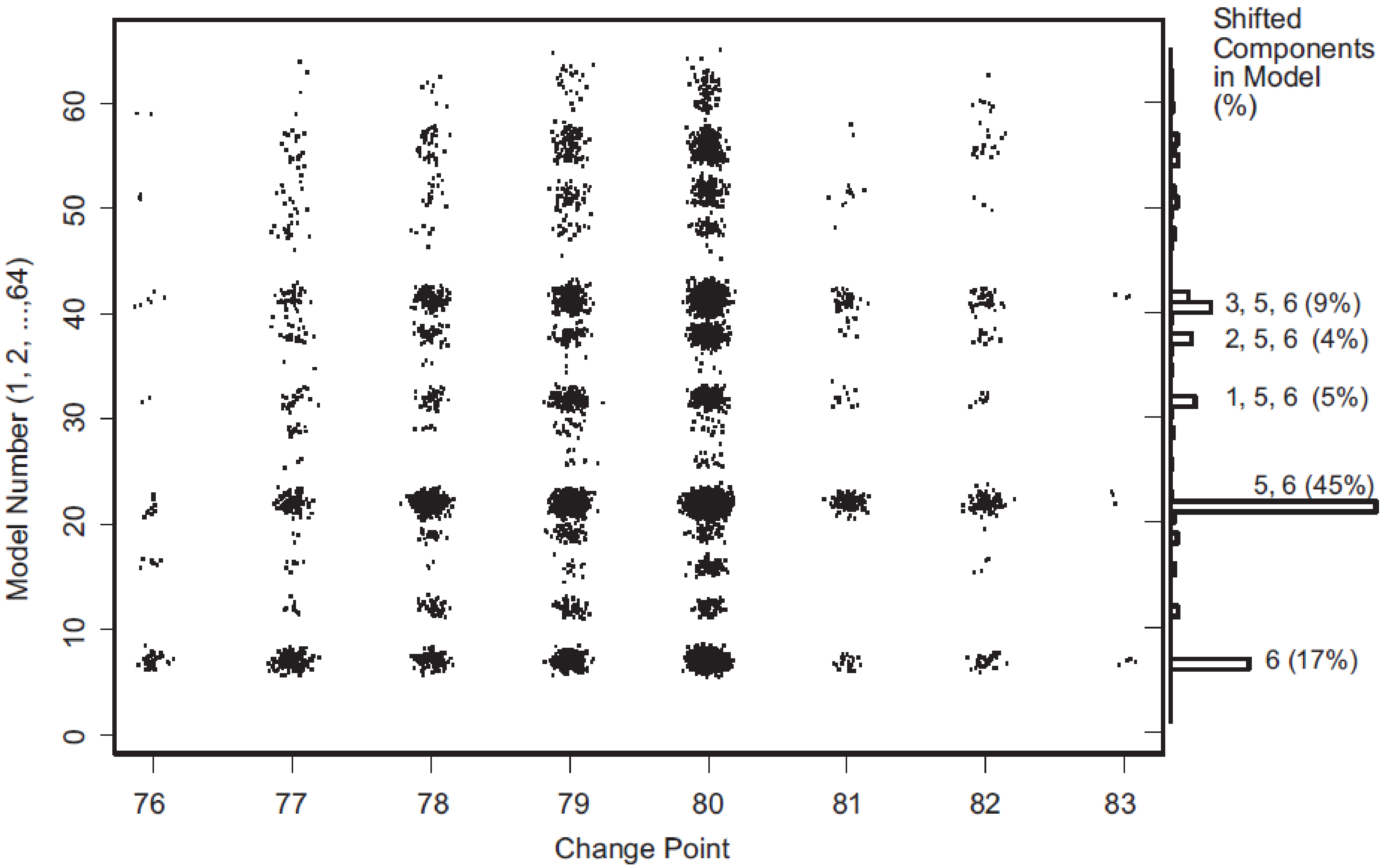
Larger model to smaller model ...

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

Current state (within model 8, $\{1,4\}$) $\tilde{\theta}_k = \left[\mu^1_{8,\tau+1}, \mu^4_{8,\tau+1}, \tau\right]$

maps to (within model 2, $\{1\}$) $\tilde{\theta}_{k'} = \left[\mu^1_{2,\tau+1}, u, \tau\right]$


Model states
model number
0
10
20
30
40
50
60
2000
4000
6000
8000
10000


Shifted
Components
in Model
(%)
3, 5, 6 (9%)
2, 5, 6 (4%)
1, 5, 6 (5%)
5, 6 (45%)
6 (17%)
Model Number (1, 2, ...,64)
0
10
20
30
40
50
60
76
77
78
79
80
81
82
83
Change Point

Model

$$\boldsymbol{X}_1, \boldsymbol{X}_2, \ldots, \boldsymbol{X}_\tau \sim N(\boldsymbol{\mu}_\tau, \Sigma)$$
$$\boldsymbol{X}_{\tau+1}, \boldsymbol{X}_{\tau+2}, \ldots, \boldsymbol{X}_N \sim N(\boldsymbol{\mu}_{\tau+1}, \Sigma)$$

Possible Assumptions

1. known: $\boldsymbol{\mu}_\tau, \Sigma$
   unknown: τ, $\boldsymbol{\mu}_{\tau+1}$
   (unrealistic, but easy to explain)
2. known: Σ
   unknown: τ, $\boldsymbol{\mu}_\tau$, $\boldsymbol{\mu}_{\tau+1}$
   (stepping stone)
3. known: nothing
   unknown: τ, $\boldsymbol{\mu}_\tau$, $\boldsymbol{\mu}_{\tau+1}, \Sigma$
   (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