A new solution to the label switching problem of MCMC outputs

Panagiotis Papastamoulis and George Iliopoulos

Department of Statistics and Insurance Science, University of Piraeus

Outline

- Independence mixture model

Outline

- Independence mixture model
- Label Switching Phenomenon

Outline

- Independence mixture model
- Label Switching Phenomenon
- A mixture representation of the posterior distribution

Outline

- Independence mixture model
- Label Switching Phenomenon
- A mixture representation of the posterior distribution
- Equivalence Classes Representative (ECR) Algorithm

Outline

- Independence mixture model
- Label Switching Phenomenon
- A mixture representation of the posterior distribution
- Equivalence Classes Representative (ECR) Algorithm
- Examples

Outline

- Independence mixture model
- Label Switching Phenomenon
- A mixture representation of the posterior distribution
- Equivalence Classes Representative (ECR) Algorithm
- Examples
- Extension: Hidden Markov models and applications

Independence mixture model

- Observed data: $\boldsymbol{y} = (y_1, \ldots, y_n)$, where $y_i \in \mathbb{R}^p$
  $Y_i \overset{\text{iid}}{\sim} \sum_{j=1}^{k} \pi_j f(y; \theta_j), \quad i = 1, \ldots, n.$

Independence mixture model

- Observed data: $\boldsymbol{y} = (y_1, \ldots, y_n)$, where $y_i \in \mathbb{R}^p$
  $Y_i \overset{\text{iid}}{\sim} \sum_{j=1}^{k} \pi_j f(y; \theta_j), \quad i = 1, \ldots, n.$
- Number of components: $k \in \mathbb{Z}^+$.

Independence mixture model

- Observed data: $\boldsymbol{y} = (y_1, \ldots, y_n)$, where $y_i \in \mathbb{R}^p$
  $Y_i \overset{\text{iid}}{\sim} \sum_{j=1}^{k} \pi_j f(y; \theta_j), \quad i = 1, \ldots, n.$
- Number of components: $k \in \mathbb{Z}^+$.
- Mixture proportions: $0 \leq \pi_j \leq 1, \forall i = 1, \ldots, k$ and $\sum_{j=1}^{k} \pi_j = 1$.

Independence mixture model

- Observed data: $\boldsymbol{y} = (y_1, \ldots, y_n)$, where $y_i \in \mathbb{R}^p$
  $Y_i \overset{\text{iid}}{\sim} \sum_{j=1}^{k} \pi_j f(y; \theta_j), \quad i = 1, \ldots, n.$
- Number of components: $k \in \mathbb{Z}^+$.
- Mixture proportions: $0 \leq \pi_j \leq 1, \forall i = 1, \ldots, k$ and $\sum_{j=1}^{k} \pi_j = 1$.
- Component specific parameters: $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_k) \in \Theta^k$

Independence mixture model

- Observed data: $\boldsymbol{y} = (y_1, \ldots, y_n)$, where $y_i \in \mathbb{R}^p$
  $Y_i \overset{\text{iid}}{\sim} \sum_{j=1}^{k} \pi_j f(y; \theta_j), \quad i = 1, \ldots, n.$
- Number of components: $k \in \mathbb{Z}^+$.
- Mixture proportions: $0 \leq \pi_j \leq 1, \forall i = 1, \ldots, k$ and $\sum_{j=1}^{k} \pi_j = 1$.
- Component specific parameters: $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_k) \in \Theta^k$
- Problem: Given k and $\boldsymbol{y}$, estimate $(\boldsymbol{\pi}, \boldsymbol{\theta})$.

Data Augmentation

- $Z_i \in \{1, \ldots, k\} : P(Z_i = j) = \pi_j$, $i = 1, \ldots, n$, independently.

Data Augmentation

- $Z_i \in \{1, \ldots, k\} : P(Z_i = j) = \pi_j$, $i = 1, \ldots, n$, independently.
- $Y_i | Z_i = j \sim f(y; \theta_j)$, $i = 1, \ldots, n$

Data Augmentation

- $Z_i \in \{1, \ldots, k\} : P(Z_i = j) = \pi_j$, $i = 1, \ldots, n$, independently.
- $Y_i | Z_i = j \sim f(y; \theta_j)$, $i = 1, \ldots, n$
- complete data: $((y_1, z_1), \ldots, (y_n, z_n))$.

Data Augmentation

- $Z_i \in \{1, \ldots, k\} : P(Z_i = j) = \pi_j$, $i = 1, \ldots, n$, independently.
- $Y_i | Z_i = j \sim f(y; \theta_j)$, $i = 1, \ldots, n$
- complete data: $((y_1, z_1), \ldots, (y_n, z_n))$.
- Don't forget: z_i's are unobserved!

Data Augmentation

- $Z_i \in \{1, \ldots, k\} : P(Z_i = j) = \pi_j$, $i = 1, \ldots, n$, independently.
- $Y_i | Z_i = j \sim f(y; \theta_j)$, $i = 1, \ldots, n$
- complete data: $((y_1, z_1), \ldots, (y_n, z_n))$.
- Don't forget: z_i's are unobserved!
- However:

$$P(Z_i = z_i | y_i, \boldsymbol{\pi}, \boldsymbol{\theta}) = \frac{\pi_{z_i} f(y_i; \theta_{z_i})}{\sum_{j=1}^{k} \pi_j f(y_i; \theta_j)}.$$

Invariance property of the posterior

- Let $\mathcal{T}_k$: set of permutations of the component indices $\{1,\ldots,k\}$.

Invariance property of the posterior

- Let $\mathcal{T}_k$: set of permutations of the component indices $\{1, \ldots, k\}$.
- Likelihood of the mixture model:
  $L(\boldsymbol{\pi}, \boldsymbol{\theta} | \boldsymbol{y}) = \prod_{i=1}^{n} \{\pi_1 f(y_i; \theta_1) + \ldots + \pi_k f(y_i; \theta_k)\}$.

Invariance property of the posterior

- Let $\mathcal{T}_k$: set of permutations of the component indices $\{1, \ldots, k\}$.
- Likelihood of the mixture model:
  $L(\boldsymbol{\pi}, \boldsymbol{\theta}|\boldsymbol{y}) = \prod_{i=1}^{n} \{\pi_1 f(y_i; \theta_1) + \ldots + \pi_k f(y_i; \theta_k)\}.$
- Invariance of the likelihood wrt permutations:
  $L(\boldsymbol{\pi}, \boldsymbol{\theta}|\boldsymbol{y}) = L(\tau(\boldsymbol{\pi}, \boldsymbol{\theta})|\boldsymbol{y}) \; \forall \tau \in \mathcal{T}_k.$

Invariance property of the posterior

- Let $\mathcal{T}_k$: set of permutations of the component indices $\{1, \ldots, k\}$.
- Likelihood of the mixture model:
  $L(\boldsymbol{\pi}, \boldsymbol{\theta}|\boldsymbol{y}) = \prod_{i=1}^{n} \{\pi_1 f(y_i; \theta_1) + \ldots + \pi_k f(y_i; \theta_k)\}.$
- Invariance of the likelihood wrt permutations:
  $L(\boldsymbol{\pi}, \boldsymbol{\theta}|\boldsymbol{y}) = L(\tau(\boldsymbol{\pi}, \boldsymbol{\theta})|\boldsymbol{y}) \ \forall \tau \in \mathcal{T}_k.$
- Suppose that the same holds for the prior
  $f(\boldsymbol{\pi}, \boldsymbol{\theta}) = f(\tau(\boldsymbol{\pi}, \boldsymbol{\theta})).$

Invariance property of the posterior

- Let $\mathcal{T}_k$: set of permutations of the component indices $\{1, \ldots, k\}$.
- Likelihood of the mixture model: $L(\boldsymbol{\pi}, \boldsymbol{\theta}|\boldsymbol{y}) = \prod_{i=1}^{n} \{\pi_1 f(y_i; \theta_1) + \ldots + \pi_k f(y_i; \theta_k)\}$.
- Invariance of the likelihood wrt permutations: $L(\boldsymbol{\pi}, \boldsymbol{\theta}|\boldsymbol{y}) = L(\tau(\boldsymbol{\pi}, \boldsymbol{\theta})|\boldsymbol{y}) \; \forall \tau \in \mathcal{T}_k$.
- Suppose that the same holds for the prior $f(\boldsymbol{\pi}, \boldsymbol{\theta}) = f(\tau(\boldsymbol{\pi}, \boldsymbol{\theta}))$.
- Posterior distribution:

$$
\begin{aligned}
f(\boldsymbol{\pi}, \boldsymbol{\theta}|\boldsymbol{y}) &\propto L(\boldsymbol{\pi}, \boldsymbol{\theta}|\boldsymbol{y}) f(\boldsymbol{\pi}, \boldsymbol{\theta}) \Rightarrow \\
f(\boldsymbol{\pi}, \boldsymbol{\theta}|\boldsymbol{y}) &= f(\tau(\boldsymbol{\pi}, \boldsymbol{\theta})|\boldsymbol{y}), \; \forall \tau \in \mathcal{T}_k.
\end{aligned}
$$

Consequences: Label Switching Problem

- Marginal posterior distribution for label $j = 1, \ldots, k$:
  $(\pi_j, \theta_j | \boldsymbol{y}) \sim \int f(\boldsymbol{\pi}, \boldsymbol{\theta} | \boldsymbol{y}) d\boldsymbol{\pi}_{[-j]} d\boldsymbol{\theta}_{[-j]}$, where
  $\boldsymbol{\theta}_{[-j]} = (\theta_1, \ldots, \theta_{j-1}, \theta_{j+1}, \ldots, \theta_k)$.

Consequences: Label Switching Problem

- Marginal posterior distribution for label $j = 1, \ldots, k$:
  $(\pi_j, \theta_j|\boldsymbol{y}) \sim \int f(\boldsymbol{\pi}, \boldsymbol{\theta}|\boldsymbol{y}) d\boldsymbol{\pi}_{[-j]} d\boldsymbol{\theta}_{[-j]}$, where
  $\boldsymbol{\theta}_{[-j]} = (\theta_1, \ldots, \theta_{j-1}, \theta_{j+1}, \ldots, \theta_k)$.
- "A-marginallia", i.e: $\forall i, j = 1, 2, \ldots, k$,
  $(\pi_1, \theta_1|\boldsymbol{y}) \equiv \ldots \equiv (\pi_k, \theta_k|\boldsymbol{y})$.

Consequences: Label Switching Problem

- Marginal posterior distribution for label $j = 1, \ldots, k$:
  $(\pi_j, \theta_j | \boldsymbol{y}) \sim \int f(\boldsymbol{\pi}, \boldsymbol{\theta} | \boldsymbol{y}) d\boldsymbol{\pi}_{[-j]} d\boldsymbol{\theta}_{[-j]}$, where
  $\boldsymbol{\theta}_{[-j]} = (\theta_1, \ldots, \theta_{j-1}, \theta_{j+1}, \ldots, \theta_k)$.
- "A-marginallia", i.e: $\forall i, j = 1, 2, \ldots, k$,
  $(\pi_1, \theta_1 | \boldsymbol{y}) \equiv \ldots \equiv (\pi_k, \theta_k | \boldsymbol{y})$.
- Previous solutions:

Consequences: Label Switching Problem

- Marginal posterior distribution for label $j = 1, \ldots, k$: $(\pi_j, \theta_j | \boldsymbol{y}) \sim \int f(\boldsymbol{\pi}, \boldsymbol{\theta} | \boldsymbol{y}) d\boldsymbol{\pi}_{[-j]} d\boldsymbol{\theta}_{[-j]}$, where $\boldsymbol{\theta}_{[-j]} = (\theta_1, \ldots, \theta_{j-1}, \theta_{j+1}, \ldots, \theta_k)$.
- "A-marginallia", i.e: $\forall i, j = 1, 2, \ldots, k$, $(\pi_1, \theta_1 | \boldsymbol{y}) \equiv \ldots \equiv (\pi_k, \theta_k | \boldsymbol{y})$.
- Previous solutions:
  1. Artificial identifiability constraints (Diebolt and Robert, 1994), e.g. $\pi_1 < \pi_2 < \ldots < \pi_k$

Consequences: Label Switching Problem

- Marginal posterior distribution for label $j = 1, \ldots, k$: $(\pi_j, \theta_j | \boldsymbol{y}) \sim \int f(\boldsymbol{\pi}, \boldsymbol{\theta} | \boldsymbol{y}) d\boldsymbol{\pi}_{[-j]} d\boldsymbol{\theta}_{[-j]}$, where $\boldsymbol{\theta}_{[-j]} = (\theta_1, \ldots, \theta_{j-1}, \theta_{j+1}, \ldots, \theta_k)$.
- “A-marginallia”, i.e: $\forall i, j = 1, 2, \ldots, k$, $(\pi_1, \theta_1 | \boldsymbol{y}) \equiv \ldots \equiv (\pi_k, \theta_k | \boldsymbol{y})$.
- Previous solutions:
  1. Artificial identifiability constraints (Diebolt and Robert, 1994), e.g. $\pi_1 < \pi_2 < \ldots < \pi_k$
  2. Pivotal Reordering algorithm (Marin et al., 2005)

Consequences: Label Switching Problem

- Marginal posterior distribution for label $j = 1, \ldots, k$: $(\pi_j, \theta_j | \boldsymbol{y}) \sim \int f(\boldsymbol{\pi}, \boldsymbol{\theta} | \boldsymbol{y}) d\boldsymbol{\pi}_{[-j]} d\boldsymbol{\theta}_{[-j]}$, where $\boldsymbol{\theta}_{[-j]} = (\theta_1, \ldots, \theta_{j-1}, \theta_{j+1}, \ldots, \theta_k)$.
- "A-marginallia", i.e: $\forall i, j = 1, 2, \ldots, k$, $(\pi_1, \theta_1 | \boldsymbol{y}) \equiv \ldots \equiv (\pi_k, \theta_k | \boldsymbol{y})$.
- Previous solutions:
  1. Artificial identifiability constraints (Diebolt and Robert, 1994), e.g. $\pi_1 < \pi_2 < \ldots < \pi_k$
  2. Pivotal Reordering algorithm (Marin et al., 2005)
  3. Kullback - Leibler divergence algorithms (Stephens, 1997)

Consequences: Label Switching Problem

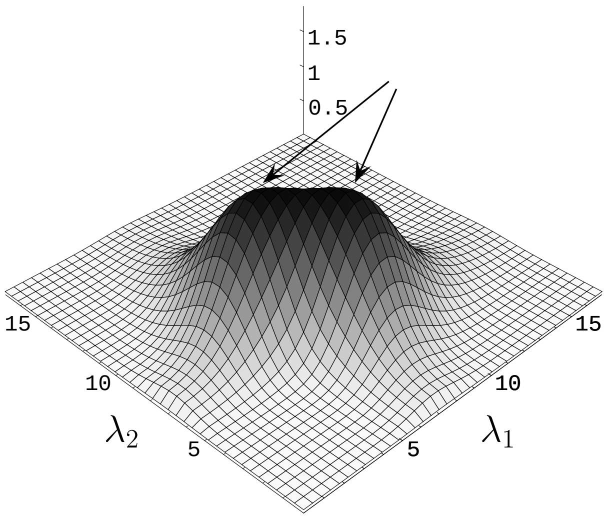
- Marginal posterior distribution for label $j = 1, \dots, k$: $(\pi_j, \theta_j | \boldsymbol{y}) \sim \int f(\boldsymbol{\pi}, \boldsymbol{\theta} | \boldsymbol{y}) d\boldsymbol{\pi}_{[-j]} d\boldsymbol{\theta}_{[-j]}$, where $\boldsymbol{\theta}_{[-j]} = (\theta_1, \dots, \theta_{j-1}, \theta_{j+1}, \dots, \theta_k)$.
- "A-marginallia", i.e: $\forall i, j = 1, 2, \dots, k$, $(\pi_1, \theta_1 | \boldsymbol{y}) \equiv \dots \equiv (\pi_k, \theta_k | \boldsymbol{y})$.
- Previous solutions:
  1. Artificial identifiability constraints (Diebolt and Robert, 1994), e.g. $\pi_1 < \pi_2 < \dots < \pi_k$
  2. Pivotal Reordering algorithm (Marin et al., 2005)
  3. Kullback - Leibler divergence algorithms (Stephens, 1997)
  4. Label invariant loss functions (Celeux et al., 2000)

Toy Example

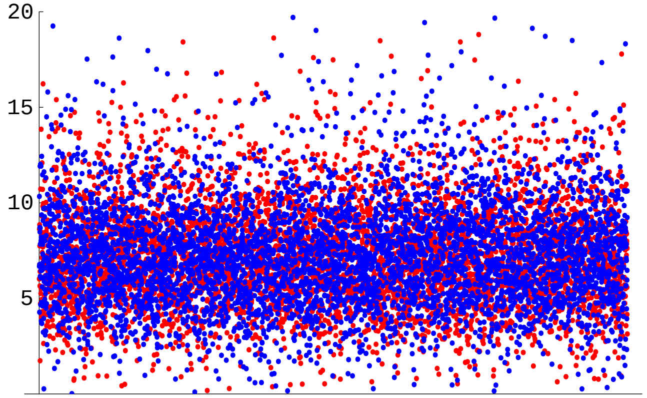
- $\boldsymbol{y} = (6, 12, 9, 4, 6) \overset{\text{iid}}{\sim} 0.5\mathcal{P}(\lambda_1) + 0.5\mathcal{P}(\lambda_2)$.

Toy Example

- $\boldsymbol{y} = (6, 12, 9, 4, 6) \overset{\text{iid}}{\sim} 0.5\mathcal{P}(\lambda_1) + 0.5\mathcal{P}(\lambda_2)$.
- (Symmetric) posterior distribution of $\lambda_1, \lambda_2 | \boldsymbol{y}$:



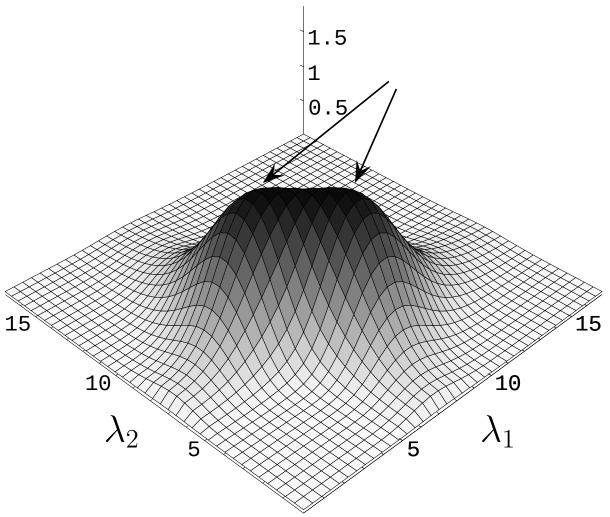
exact evaluation



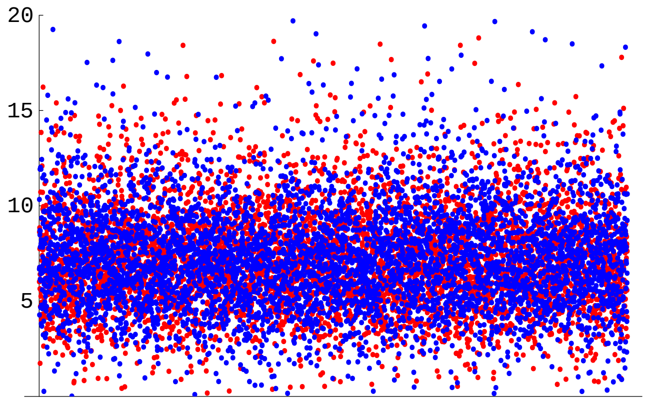
Gibbs Sampler

Toy Example

- $\boldsymbol{y} = (6, 12, 9, 4, 6) \overset{\text{iid}}{\sim} 0.5\mathcal{P}(\lambda_1) + 0.5\mathcal{P}(\lambda_2)$.
- (Symmetric) posterior distribution of $\lambda_1, \lambda_2 | \boldsymbol{y}$:



exact evaluation



Gibbs Sampler

- How we will separate the symmetric areas?

A mixture representation of the posterior distribution

Definition 1: Let $\tau = (t_1, \ldots, t_k) \in \mathcal{T}_k$, be a permutation of the index set. The corresponding relabelling $\tau \boldsymbol{z}$ of the allocation vector $\boldsymbol{z} = (z_1, \ldots, z_n) \in \mathcal{Z} := \{1, \ldots, k\}^n$ is given by $\tau \boldsymbol{z} = (t_{z_1}, \ldots, t_{z_n}) \in \mathcal{Z}$.

A mixture representation of the posterior distribution

Definition 1: Let $\tau = (t_1, \ldots, t_k) \in \mathcal{T}_k$, be a permutation of the index set. The corresponding relabelling $\tau \boldsymbol{z}$ of the allocation vector $\boldsymbol{z} = (z_1, \ldots, z_n) \in \mathcal{Z} := \{1, \ldots, k\}^n$ is given by $\tau \boldsymbol{z} = (t_{z_1}, \ldots, t_{z_n}) \in \mathcal{Z}$.

Example: Let $n = 10, k = 3, \boldsymbol{z} = \{2, 1, 2, 2, 3, 1, 2, 3, 3, 3\}$ and $\tau = \{2, 3, 1\}$. Then, $\tau \boldsymbol{z} = \{3, 2, 3, 3, 1, 2, 3, 1, 1, 1\}$.

A mixture representation of the posterior distribution

Definition 1: Let $\tau = (t_1, \ldots, t_k) \in \mathcal{T}_k$, be a permutation of the index set. The corresponding relabelling $\tau\boldsymbol{z}$ of the allocation vector $\boldsymbol{z} = (z_1, \ldots, z_n) \in \mathcal{Z} := \{1, \ldots, k\}^n$ is given by $\tau\boldsymbol{z} = (t_{z_1}, \ldots, t_{z_n}) \in \mathcal{Z}$.

Example: Let $n = 10, k = 3, \boldsymbol{z} = \{2, 1, 2, 2, 3, 1, 2, 3, 3, 3\}$ and $\tau = \{2, 3, 1\}$. Then, $\tau\boldsymbol{z} = \{3, 2, 3, 3, 1, 2, 3, 1, 1, 1\}$.

Definition 2: Two allocation vectors $\boldsymbol{z}_1, \boldsymbol{z}_2 \in \mathcal{Z}$ will be said to be equivalent if there exists $\tau \in \mathcal{T}_k$ such that $\boldsymbol{z}_1 = \tau\boldsymbol{z}_2$.

A mixture representation of the posterior distribution

Definition 1: Let $\tau = (t_1, \ldots, t_k) \in \mathcal{T}_k$, be a permutation of the index set. The corresponding relabelling $\tau\boldsymbol{z}$ of the allocation vector $\boldsymbol{z} = (z_1, \ldots, z_n) \in \mathcal{Z} := \{1, \ldots, k\}^n$ is given by $\tau\boldsymbol{z} = (t_{z_1}, \ldots, t_{z_n}) \in \mathcal{Z}$.

Example: Let $n = 10, k = 3, \boldsymbol{z} = \{2, 1, 2, 2, 3, 1, 2, 3, 3, 3\}$ and $\tau = \{2, 3, 1\}$. Then, $\tau\boldsymbol{z} = \{3, 2, 3, 3, 1, 2, 3, 1, 1, 1\}$.

Definition 2: Two allocation vectors $\boldsymbol{z}_1, \boldsymbol{z}_2 \in \mathcal{Z}$ will be said to be equivalent if there exists $\tau \in \mathcal{T}_k$ such that $\boldsymbol{z}_1 = \tau\boldsymbol{z}_2$.

- Let $\Xi_{\boldsymbol{z}} = \{\tau\boldsymbol{z} : \tau \in \mathcal{T}_k\}$ be the equivalence class of $\boldsymbol{z}$ implied by the equivalence relation.

A mixture representation of the posterior distribution

- Consider now an arbitrary set $\mathcal{Z}_0$, consisting of exactly one representative from each equivalence class and let

$$f_{\mathcal{Z}_0}(\boldsymbol{\pi}, \boldsymbol{\theta}|\boldsymbol{y}) := \sum_{\boldsymbol{z} \in \mathcal{Z}_0} \frac{k!\, w(\boldsymbol{z}|\boldsymbol{y})}{(k - k_0(\boldsymbol{z}))!} f(\boldsymbol{\pi}, \boldsymbol{\theta}|\boldsymbol{y}, \boldsymbol{z}),$$

  where $w(\boldsymbol{z}|\boldsymbol{y})$ and $k_0(\boldsymbol{z})$: posterior weight and number of nonempty components for a given allocation vector $\boldsymbol{z}$, respectively.

A mixture representation of the posterior distribution

- Consider now an arbitrary set $\mathcal{Z}_0$, consisting of exactly one representative from each equivalence class and let

$$f_{\mathcal{Z}_0}(\boldsymbol{\pi}, \boldsymbol{\theta}|\boldsymbol{y}) := \sum_{\boldsymbol{z} \in \mathcal{Z}_0} \frac{k!\, w(\boldsymbol{z}|\boldsymbol{y})}{(k - k_0(\boldsymbol{z}))!} f(\boldsymbol{\pi}, \boldsymbol{\theta}|\boldsymbol{y}, \boldsymbol{z}),$$

  where $w(\boldsymbol{z}|\boldsymbol{y})$ and $k_0(\boldsymbol{z})$: posterior weight and number of nonempty components for a given allocation vector $\boldsymbol{z}$, respectively.
- $f_{\mathcal{Z}_0}$ is a distribution for $(\boldsymbol{\pi}, \boldsymbol{\theta})$ with the same support as the posterior.

A mixture representation of the posterior distribution

- Consider now an arbitrary set $\mathcal{Z}_0$, consisting of exactly one representative from each equivalence class and let

$$f_{\mathcal{Z}_0}(\boldsymbol{\pi}, \boldsymbol{\theta}|\boldsymbol{y}) := \sum_{\boldsymbol{z} \in \mathcal{Z}_0} \frac{k! \, w(\boldsymbol{z}|\boldsymbol{y})}{(k - k_0(\boldsymbol{z}))!} f(\boldsymbol{\pi}, \boldsymbol{\theta}|\boldsymbol{y}, \boldsymbol{z}),$$

  where $w(\boldsymbol{z}|\boldsymbol{y})$ and $k_0(\boldsymbol{z})$: posterior weight and number of nonempty components for a given allocation vector $\boldsymbol{z}$, respectively.
- $f_{\mathcal{Z}_0}$ is a distribution for $(\boldsymbol{\pi}, \boldsymbol{\theta})$ with the same support as the posterior.
- $f_{\mathcal{Z}_0}$ non-symmetric.

A mixture representation of the posterior distribution

Theorem: The posterior distribution of $(\boldsymbol{\pi}, \boldsymbol{\theta})$ can be expressed as

$$f(\boldsymbol{\pi}, \boldsymbol{\theta}|\boldsymbol{y}) = \frac{1}{k!} \sum_{\tau \in \mathcal{T}_k} f_{\mathcal{Z}_0}\left(\tau(\boldsymbol{\pi}, \boldsymbol{\theta})|\boldsymbol{y}\right)$$

for any choice of the equivalence classes representatives $\mathcal{Z}_0$.

A mixture representation of the posterior distribution

Theorem: The posterior distribution of $(\boldsymbol{\pi}, \boldsymbol{\theta})$ can be expressed as

$$f(\boldsymbol{\pi}, \boldsymbol{\theta}|\boldsymbol{y}) = \frac{1}{k!} \sum_{\tau \in \mathcal{T}_k} f_{\mathcal{Z}_0} \left(\tau(\boldsymbol{\pi}, \boldsymbol{\theta})|\boldsymbol{y}\right)$$

for any choice of the equivalence classes representatives $\mathcal{Z}_0$.

In a few words: Posterior symmetry = mixture of non-symmetric distributions

A mixture representation of the posterior distribution

Theorem: The posterior distribution of $(\boldsymbol{\pi}, \boldsymbol{\theta})$ can be expressed as

$$f(\boldsymbol{\pi}, \boldsymbol{\theta}|\boldsymbol{y}) = \frac{1}{k!} \sum_{\tau \in \mathcal{T}_k} f_{\mathcal{Z}_0}\left(\tau(\boldsymbol{\pi}, \boldsymbol{\theta})|\boldsymbol{y}\right)$$

for any choice of the equivalence classes representatives $\mathcal{Z}_0$.

In a few words: Posterior symmetry = mixture of non-symmetric distributions

Consequently: Work with the non symmetric components!

Convergence Result

- Set $\mathcal{T}_{\boldsymbol{z}} := \{\tau \in \mathcal{T}_k : \tau \boldsymbol{z} \in \mathcal{Z}_0\}$.

Convergence Result

- Set $\mathcal{T}_{\boldsymbol{z}} := \{\tau \in \mathcal{T}_k : \tau \boldsymbol{z} \in \mathcal{Z}_0\}$.
- Let $(\boldsymbol{z}^{(i)}, (\boldsymbol{\pi}, \boldsymbol{\theta})^{(i)})$, $i = 1, 2, \ldots$ be a Markov chain with limit distribution $f(\boldsymbol{z}, \boldsymbol{\pi}, \boldsymbol{\theta} | \boldsymbol{y})$. For all i let τ_i be uniformly distributed on $\mathcal{T}_{\boldsymbol{z}^{(i)}}$. Then the limit distribution of the sequence $\tau_i^{-1}(\boldsymbol{\pi}, \boldsymbol{\theta})^{(i)}$ is $f_{\mathcal{Z}_0}(\boldsymbol{\pi}, \boldsymbol{\theta} | \boldsymbol{y})$.

The Equivalence Classes Representatives (ECR) Algorithm

- Determine $\mathcal{Z}_0$ and obtain a simulated output $(\boldsymbol{z}, \boldsymbol{\pi}, \boldsymbol{\theta})^{(i)}$, $i = 1, \ldots, M$, with target distribution $f(\boldsymbol{z}, \boldsymbol{\pi}, \boldsymbol{\theta} | \boldsymbol{y})$.

The Equivalence Classes Representatives (ECR) Algorithm

- Determine $\mathcal{Z}_0$ and obtain a simulated output $(\boldsymbol{z}, \boldsymbol{\pi}, \boldsymbol{\theta})^{(i)}$, $i = 1, \ldots, M$, with target distribution $f(\boldsymbol{z}, \boldsymbol{\pi}, \boldsymbol{\theta} | \boldsymbol{y})$.
- For $i = 1, \ldots, M$:

The Equivalence Classes Representatives (ECR) Algorithm

- Determine $\mathcal{Z}_0$ and obtain a simulated output $(\boldsymbol{z}, \boldsymbol{\pi}, \boldsymbol{\theta})^{(i)}$, $i = 1, \ldots, M$, with target distribution $f(\boldsymbol{z}, \boldsymbol{\pi}, \boldsymbol{\theta} | \boldsymbol{y})$.
- For $i = 1, \ldots, M$:
  1. Select randomly some permutation $\tau_i \in \mathcal{T}_{\boldsymbol{z}^{(i)}}$.

The Equivalence Classes Representatives (ECR) Algorithm

- Determine $\mathcal{Z}_0$ and obtain a simulated output $(\boldsymbol{z}, \boldsymbol{\pi}, \boldsymbol{\theta})^{(i)}$, $i = 1, \ldots, M$, with target distribution $f(\boldsymbol{z}, \boldsymbol{\pi}, \boldsymbol{\theta}|\boldsymbol{y})$.
- For $i = 1, \ldots, M$:
  1. Select randomly some permutation $\tau_i \in \mathcal{T}_{\boldsymbol{z}^{(i)}}$.
  2. Set $(\boldsymbol{z}', \boldsymbol{\pi}', \boldsymbol{\theta}')^{(i)} = \left(\tau_i \boldsymbol{z}^{(i)}, \tau_i^{-1}(\boldsymbol{\pi}, \boldsymbol{\theta}^{(i)})\right)$.

The Equivalence Classes Representatives (ECR) Algorithm

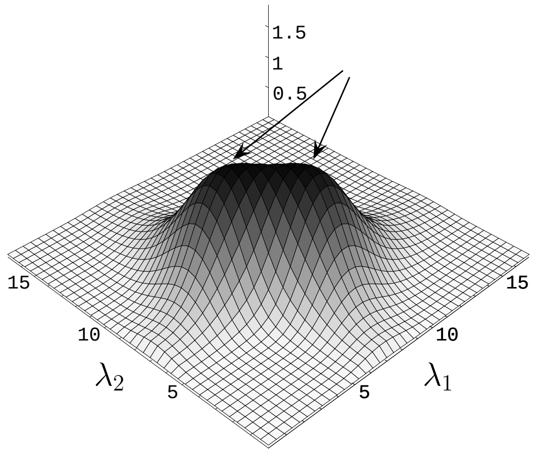
- Determine $\mathcal{Z}_0$ and obtain a simulated output $(\boldsymbol{z}, \boldsymbol{\pi}, \boldsymbol{\theta})^{(i)}$, $i = 1, \ldots, M$, with target distribution $f(\boldsymbol{z}, \boldsymbol{\pi}, \boldsymbol{\theta} | \boldsymbol{y})$.
- For $i = 1, \ldots, M$:
  1. Select randomly some permutation $\tau_i \in \mathcal{T}_{\boldsymbol{z}^{(i)}}$.
  2. Set $(\boldsymbol{z}', \boldsymbol{\pi}', \boldsymbol{\theta}')^{(i)} = \left(\tau_i \boldsymbol{z}^{(i)}, \tau_i^{-1}(\boldsymbol{\pi}, \boldsymbol{\theta}^{(i)})\right)$.
- Use the reordered values $(\boldsymbol{\pi}', \boldsymbol{\theta}')^{(i)}$, $i = 1, \ldots, M$, in order to approximate the posterior distributions of the weights and of the component specific parameters.

Toy Example revisited

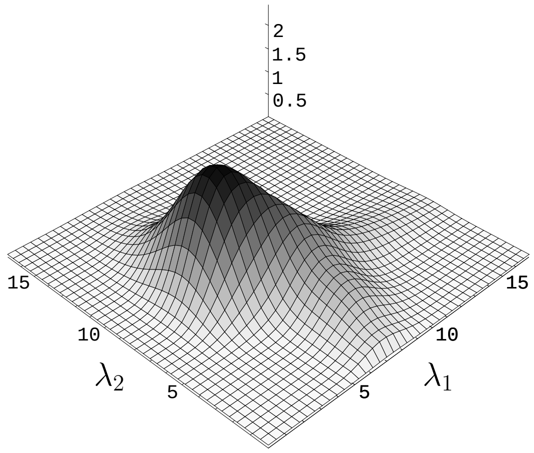
- $\boldsymbol{y} = (6, 12, 9, 4, 6) \overset{\text{iid}}{\sim} 0.5\mathcal{P}(\lambda_1) + 0.5\mathcal{P}(\lambda_2)$.

Toy Example revisited

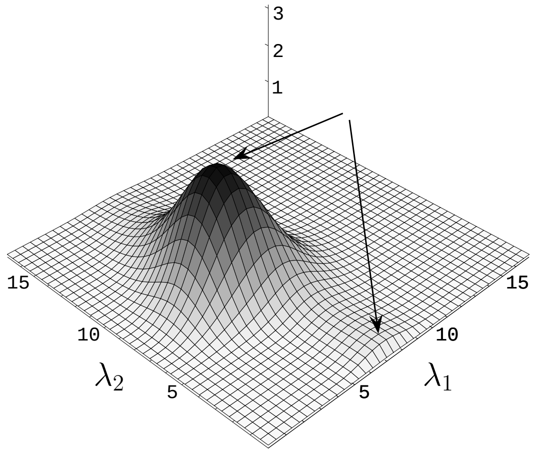
- $\boldsymbol{y} = (6, 12, 9, 4, 6) \overset{\text{iid}}{\sim} 0.5\mathcal{P}(\lambda_1) + 0.5\mathcal{P}(\lambda_2)$.
- (Symmetric) posterior distribution of $\lambda_1, \lambda_2 | \boldsymbol{y}$:



$f(\lambda_1, \lambda_2 | \boldsymbol{y})$



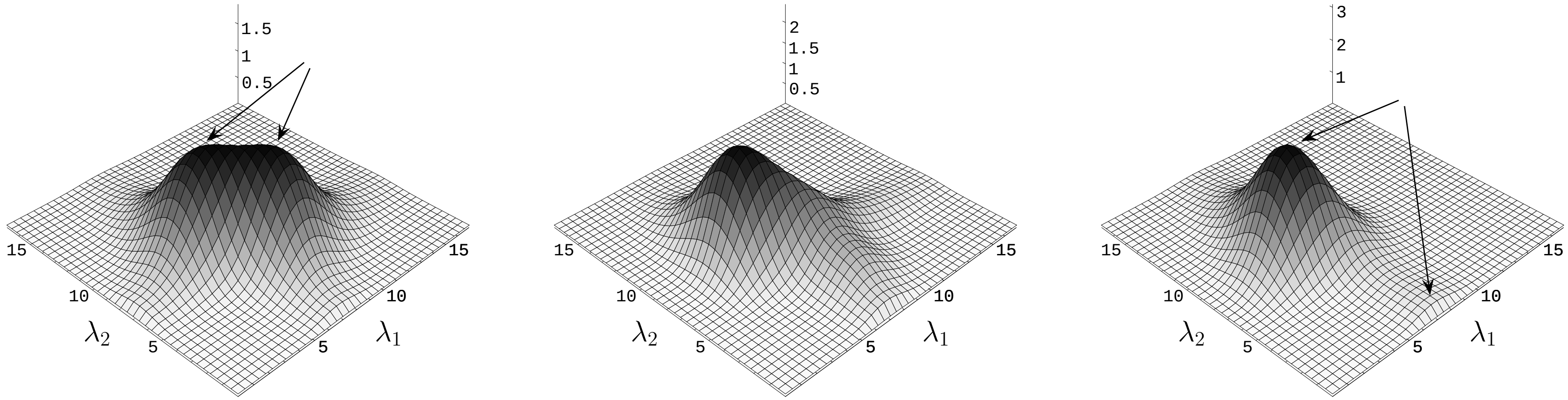
$f_{\mathcal{Z}_0'}(\lambda_1, \lambda_2 | \boldsymbol{y})$



$f_{\mathcal{Z}_0}(\lambda_1, \lambda_2 | \boldsymbol{y})$

Toy Example revisited

- $\boldsymbol{y} = (6, 12, 9, 4, 6) \overset{\text{iid}}{\sim} 0.5\mathcal{P}(\lambda_1) + 0.5\mathcal{P}(\lambda_2)$.
- (Symmetric) posterior distribution of $\lambda_1, \lambda_2 | \boldsymbol{y}$:

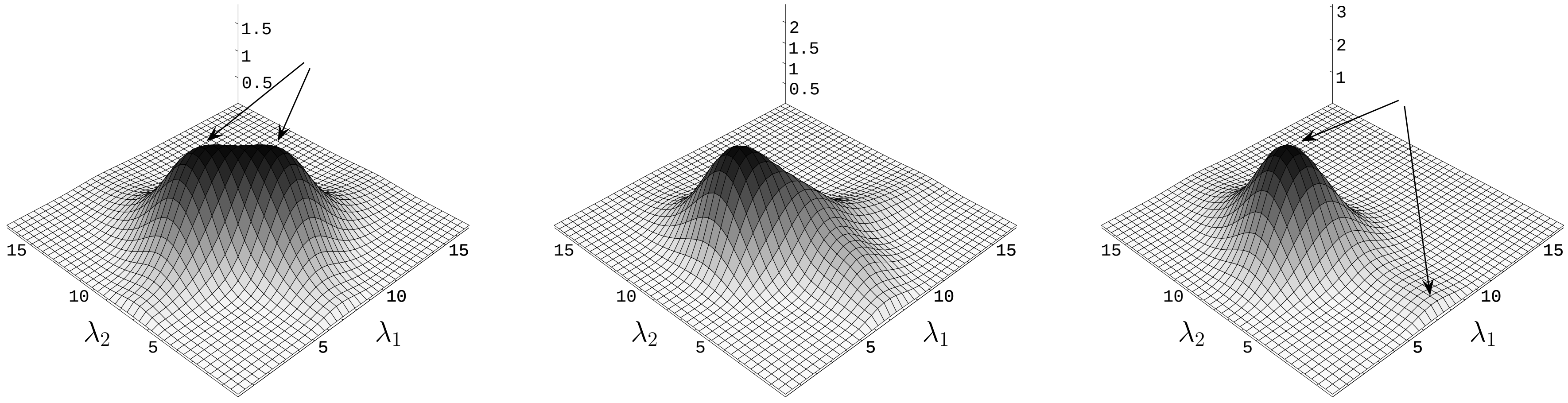


$f(\lambda_1, \lambda_2|\boldsymbol{y})$ $\qquad f_{\mathcal{Z}'_0}(\lambda_1, \lambda_2|\boldsymbol{y})$ $\qquad f_{\mathcal{Z}_0}(\lambda_1, \lambda_2|\boldsymbol{y})$

- In $\mathcal{Z}'_0$: random selection of representatives.

Toy Example revisited

- $\boldsymbol{y} = (6, 12, 9, 4, 6) \overset{\text{iid}}{\sim} 0.5\mathcal{P}(\lambda_1) + 0.5\mathcal{P}(\lambda_2)$.
- (Symmetric) posterior distribution of $\lambda_1, \lambda_2 | \boldsymbol{y}$:



$f(\lambda_1, \lambda_2 | \boldsymbol{y})$ $\quad$ $f_{\mathcal{Z}_0'}(\lambda_1, \lambda_2 | \boldsymbol{y})$ $\quad$ $f_{\mathcal{Z}_0}(\lambda_1, \lambda_2 | \boldsymbol{y})$

- In $\mathcal{Z}_0'$: random selection of representatives.
- In $\mathcal{Z}_0$: ?

Selecting the representatives

Definition: The S similarity measure of two allocation vectors $\boldsymbol{z}_1 = (z_{11}, \ldots, z_{1n})$ and $\boldsymbol{z}_2 = (z_{21}, \ldots, z_{2n})$ is defined by $S(\boldsymbol{z}_1, \boldsymbol{z}_2) := \sum_{i=1}^{n} I(z_{1i} = z_{2i})$, where $I(A)$ is the indicator function of A.

Selecting the representatives

Definition: The S similarity measure of two allocation vectors $\boldsymbol{z}_1 = (z_{11}, \ldots, z_{1n})$ and $\boldsymbol{z}_2 = (z_{21}, \ldots, z_{2n})$ is defined by $S(\boldsymbol{z}_1, \boldsymbol{z}_2) := \sum_{i=1}^{n} I(z_{1i} = z_{2i})$, where $I(A)$ is the indicator function of A.

- Find the Maximum A Posteriori (MAP) estimate: $(\hat{\boldsymbol{z}}, \hat{\boldsymbol{\pi}}, \hat{\boldsymbol{\theta}})^{(\mathrm{MAP})} = \arg\max_{1 \le i \le M} f((\boldsymbol{z}, \boldsymbol{\pi}, \boldsymbol{\theta})^{(i)} | \boldsymbol{y})$.

Selecting the representatives

Definition: The S similarity measure of two allocation vectors $\boldsymbol{z}_1 = (z_{11}, \ldots, z_{1n})$ and $\boldsymbol{z}_2 = (z_{21}, \ldots, z_{2n})$ is defined by $S(\boldsymbol{z}_1, \boldsymbol{z}_2) := \sum_{i=1}^{n} I(z_{1i} = z_{2i})$, where $I(A)$ is the indicator function of A.

- Find the Maximum A Posteriori (MAP) estimate: $(\hat{\boldsymbol{z}}, \hat{\boldsymbol{\pi}}, \hat{\boldsymbol{\theta}})^{(\mathrm{MAP})} = \arg\max_{1 \le i \le M} f((\boldsymbol{z}, \boldsymbol{\pi}, \boldsymbol{\theta})^{(i)} | \boldsymbol{y})$.
- $\tau(\Xi_{\boldsymbol{z}}) = \arg\max_{\tau \in \mathcal{T}_k} S(\tau \boldsymbol{z}, \hat{\boldsymbol{z}}^{\mathrm{MAP}})$.

Selecting the representatives

Definition: The S similarity measure of two allocation vectors $\boldsymbol{z}_1 = (z_{11}, \ldots, z_{1n})$ and $\boldsymbol{z}_2 = (z_{21}, \ldots, z_{2n})$ is defined by $S(\boldsymbol{z}_1, \boldsymbol{z}_2) := \sum_{i=1}^{n} I(z_{1i} = z_{2i})$, where $I(A)$ is the indicator function of A.

- Find the Maximum A Posteriori (MAP) estimate: $(\hat{\boldsymbol{z}}, \hat{\boldsymbol{\pi}}, \hat{\boldsymbol{\theta}})^{(\mathrm{MAP})} = \arg\max_{1 \le i \le M} f((\boldsymbol{z}, \boldsymbol{\pi}, \boldsymbol{\theta})^{(i)} | \boldsymbol{y})$.
- $\tau(\Xi_{\boldsymbol{z}}) = \arg\max_{\tau \in \mathcal{T}_k} S(\tau \boldsymbol{z}, \hat{\boldsymbol{z}}^{\mathrm{MAP}})$.
- $\mathcal{Z}_0 = \cup \{\tau(\Xi_{\boldsymbol{z}})\boldsymbol{z};\ \boldsymbol{z} \in \mathcal{Z}\}$.

Example: Reordering a single iteration

- Let $n = 10$, $k = 3$, $\boldsymbol{z}^{(i)} = (1, 2, 3, 3, 1, 2, 1, 3, 3, 2)$, $\boldsymbol{\pi}^{(i)} = (\pi_1^{(i)}, \pi_2^{(i)}, \pi_3^{(i)})$ and $\boldsymbol{\theta}^{(i)} = (\theta_1^{(i)}, \theta_2^{(i)}, \theta_3^{(i)})$.

Example: Reordering a single iteration

- Let $n = 10$, $k = 3$, $\boldsymbol{z}^{(i)} = (1, 2, 3, 3, 1, 2, 1, 3, 3, 2)$, $\boldsymbol{\pi}^{(i)} = (\pi_1^{(i)}, \pi_2^{(i)}, \pi_3^{(i)})$ and $\boldsymbol{\theta}^{(i)} = (\theta_1^{(i)}, \theta_2^{(i)}, \theta_3^{(i)})$.
- Suppose that $\hat{\boldsymbol{z}}^{\mathrm{MAP}} = (3, 3, 1, 1, 2, 3, 2, 1, 1, 2)$.

Example: Reordering a single iteration

- Let $n = 10$, $k = 3$, $\boldsymbol{z}^{(i)} = (1, 2, 3, 3, 1, 2, 1, 3, 3, 2)$, $\boldsymbol{\pi}^{(i)} = (\pi_1^{(i)}, \pi_2^{(i)}, \pi_3^{(i)})$ and $\boldsymbol{\theta}^{(i)} = (\theta_1^{(i)}, \theta_2^{(i)}, \theta_3^{(i)})$.
- Suppose that $\hat{\boldsymbol{z}}^{\mathrm{MAP}} = (3, 3, 1, 1, 2, 3, 2, 1, 1, 2)$.
- Find the optimal permutation:

Example: Reordering a single iteration

- Let $n = 10$, $k = 3$, $\boldsymbol{z}^{(i)} = (1, 2, 3, 3, 1, 2, 1, 3, 3, 2)$, $\boldsymbol{\pi}^{(i)} = (\pi_1^{(i)}, \pi_2^{(i)}, \pi_3^{(i)})$ and $\boldsymbol{\theta}^{(i)} = (\theta_1^{(i)}, \theta_2^{(i)}, \theta_3^{(i)})$.
- Suppose that $\hat{\boldsymbol{z}}^{\mathrm{MAP}} = (3, 3, 1, 1, 2, 3, 2, 1, 1, 2)$.
- Find the optimal permutation:
  1. $\tau = \{1, 2, 3\} \Rightarrow \tau\boldsymbol{z}^{(i)} = (1, 2, 3, 3, 1, 2, 1, 3, 3, 2)$ and $S(\tau\boldsymbol{z}^{(i)}, \hat{\boldsymbol{z}}^{\mathrm{MAP}}) = 1$.

Example: Reordering a single iteration

- Let $n = 10$, $k = 3$, $\boldsymbol{z}^{(i)} = (1, 2, 3, 3, 1, 2, 1, 3, 3, 2)$, $\boldsymbol{\pi}^{(i)} = (\pi_1^{(i)}, \pi_2^{(i)}, \pi_3^{(i)})$ and $\boldsymbol{\theta}^{(i)} = (\theta_1^{(i)}, \theta_2^{(i)}, \theta_3^{(i)})$.
- Suppose that $\hat{\boldsymbol{z}}^{\mathrm{MAP}} = (3, 3, 1, 1, 2, 3, 2, 1, 1, 2)$.
- Find the optimal permutation:
  1. $\tau = \{1, 2, 3\} \Rightarrow \tau\boldsymbol{z}^{(i)} = (1, 2, 3, 3, 1, 2, 1, 3, 3, 2)$ and $S(\tau\boldsymbol{z}^{(i)}, \hat{\boldsymbol{z}}^{\mathrm{MAP}}) = 1$.
  2. $\tau = \{1, 3, 2\} \Rightarrow \tau\boldsymbol{z}^{(i)} = (1, 3, 2, 2, 1, 3, 1, 2, 2, 3)$ and $S(\tau\boldsymbol{z}^{(i)}, \hat{\boldsymbol{z}}^{\mathrm{MAP}}) = 2$.

Example: Reordering a single iteration

- Let $n = 10$, $k = 3$, $\boldsymbol{z}^{(i)} = (1, 2, 3, 3, 1, 2, 1, 3, 3, 2)$, $\boldsymbol{\pi}^{(i)} = (\pi_1^{(i)}, \pi_2^{(i)}, \pi_3^{(i)})$ and $\boldsymbol{\theta}^{(i)} = (\theta_1^{(i)}, \theta_2^{(i)}, \theta_3^{(i)})$.
- Suppose that $\hat{\boldsymbol{z}}^{\mathrm{MAP}} = (3, 3, 1, 1, 2, 3, 2, 1, 1, 2)$.
- Find the optimal permutation:
  1. $\tau = \{1, 2, 3\} \Rightarrow \tau\boldsymbol{z}^{(i)} = (1, 2, 3, 3, 1, 2, 1, 3, 3, 2)$ and $S(\tau\boldsymbol{z}^{(i)}, \hat{\boldsymbol{z}}^{\mathrm{MAP}}) = 1$.
  2. $\tau = \{1, 3, 2\} \Rightarrow \tau\boldsymbol{z}^{(i)} = (1, 3, 2, 2, 1, 3, 1, 2, 2, 3)$ and $S(\tau\boldsymbol{z}^{(i)}, \hat{\boldsymbol{z}}^{\mathrm{MAP}}) = 2$.
  3. $\tau = \{2, 1, 3\} \Rightarrow \tau\boldsymbol{z}^{(i)} = (2, 1, 3, 3, 2, 1, 2, 3, 3, 1)$ and $S(\tau\boldsymbol{z}^{(i)}, \hat{\boldsymbol{z}}^{\mathrm{MAP}}) = 2$.

Example: Reordering a single iteration

- Let $n = 10$, $k = 3$, $\boldsymbol{z}^{(i)} = (1, 2, 3, 3, 1, 2, 1, 3, 3, 2)$, $\boldsymbol{\pi}^{(i)} = (\pi_1^{(i)}, \pi_2^{(i)}, \pi_3^{(i)})$ and $\boldsymbol{\theta}^{(i)} = (\theta_1^{(i)}, \theta_2^{(i)}, \theta_3^{(i)})$.
- Suppose that $\hat{\boldsymbol{z}}^{\mathrm{MAP}} = (3, 3, 1, 1, 2, 3, 2, 1, 1, 2)$.
- Find the optimal permutation:
  1. $\tau = \{1, 2, 3\} \Rightarrow \tau\boldsymbol{z}^{(i)} = (1, 2, 3, 3, 1, 2, 1, 3, 3, 2)$ and $S(\tau\boldsymbol{z}^{(i)}, \hat{\boldsymbol{z}}^{\mathrm{MAP}}) = 1$.
  2. $\tau = \{1, 3, 2\} \Rightarrow \tau\boldsymbol{z}^{(i)} = (1, 3, 2, 2, 1, 3, 1, 2, 2, 3)$ and $S(\tau\boldsymbol{z}^{(i)}, \hat{\boldsymbol{z}}^{\mathrm{MAP}}) = 2$.
  3. $\tau = \{2, 1, 3\} \Rightarrow \tau\boldsymbol{z}^{(i)} = (2, 1, 3, 3, 2, 1, 2, 3, 3, 1)$ and $S(\tau\boldsymbol{z}^{(i)}, \hat{\boldsymbol{z}}^{\mathrm{MAP}}) = 2$.
  4. $\tau = \{2, 3, 1\} \Rightarrow \tau\boldsymbol{z}^{(i)} = (2, 3, 1, 1, 2, 3, 2, 1, 1, 3)$ and $S(\tau\boldsymbol{z}^{(i)}, \hat{\boldsymbol{z}}^{\mathrm{MAP}}) = 8$.

Example: Reordering a single iteration

- Let $n = 10$, $k = 3$, $\boldsymbol{z}^{(i)} = (1,2,3,3,1,2,1,3,3,2)$, $\boldsymbol{\pi}^{(i)} = (\pi_1^{(i)}, \pi_2^{(i)}, \pi_3^{(i)})$ and $\boldsymbol{\theta}^{(i)} = (\theta_1^{(i)}, \theta_2^{(i)}, \theta_3^{(i)})$.
- Suppose that $\hat{\boldsymbol{z}}^{\mathrm{MAP}} = (3,3,1,1,2,3,2,1,1,2)$.
- Find the optimal permutation:
  1. $\tau = \{1,2,3\} \Rightarrow \tau\boldsymbol{z}^{(i)} = (1,2,3,3,1,2,1,3,3,2)$ and $S(\tau\boldsymbol{z}^{(i)}, \hat{\boldsymbol{z}}^{\mathrm{MAP}}) = 1$.
  2. $\tau = \{1,3,2\} \Rightarrow \tau\boldsymbol{z}^{(i)} = (1,3,2,2,1,3,1,2,2,3)$ and $S(\tau\boldsymbol{z}^{(i)}, \hat{\boldsymbol{z}}^{\mathrm{MAP}}) = 2$.
  3. $\tau = \{2,1,3\} \Rightarrow \tau\boldsymbol{z}^{(i)} = (2,1,3,3,2,1,2,3,3,1)$ and $S(\tau\boldsymbol{z}^{(i)}, \hat{\boldsymbol{z}}^{\mathrm{MAP}}) = 2$.
  4. $\tau = \{2,3,1\} \Rightarrow \tau\boldsymbol{z}^{(i)} = (2,3,1,1,2,3,2,1,1,3)$ and $S(\tau\boldsymbol{z}^{(i)}, \hat{\boldsymbol{z}}^{\mathrm{MAP}}) = 8$.
  5. $\tau = \{3,1,2\} \Rightarrow \tau\boldsymbol{z}^{(i)} = (3,1,2,2,3,1,3,2,2,1)$ and $S(\tau\boldsymbol{z}^{(i)}, \hat{\boldsymbol{z}}^{\mathrm{MAP}}) = 1$.

Example: Reordering a single iteration

- Let $n = 10$, $k = 3$, $\boldsymbol{z}^{(i)} = (1, 2, 3, 3, 1, 2, 1, 3, 3, 2)$, $\boldsymbol{\pi}^{(i)} = (\pi_1^{(i)}, \pi_2^{(i)}, \pi_3^{(i)})$ and $\boldsymbol{\theta}^{(i)} = (\theta_1^{(i)}, \theta_2^{(i)}, \theta_3^{(i)})$.
- Suppose that $\hat{\boldsymbol{z}}^{\mathrm{MAP}} = (3, 3, 1, 1, 2, 3, 2, 1, 1, 2)$.
- Find the optimal permutation:
  1. $\tau = \{1, 2, 3\} \Rightarrow \tau\boldsymbol{z}^{(i)} = (1, 2, 3, 3, 1, 2, 1, 3, 3, 2)$ and $S(\tau\boldsymbol{z}^{(i)}, \hat{\boldsymbol{z}}^{\mathrm{MAP}}) = 1$.
  2. $\tau = \{1, 3, 2\} \Rightarrow \tau\boldsymbol{z}^{(i)} = (1, 3, 2, 2, 1, 3, 1, 2, 2, 3)$ and $S(\tau\boldsymbol{z}^{(i)}, \hat{\boldsymbol{z}}^{\mathrm{MAP}}) = 2$.
  3. $\tau = \{2, 1, 3\} \Rightarrow \tau\boldsymbol{z}^{(i)} = (2, 1, 3, 3, 2, 1, 2, 3, 3, 1)$ and $S(\tau\boldsymbol{z}^{(i)}, \hat{\boldsymbol{z}}^{\mathrm{MAP}}) = 2$.
  4. $\tau = \{2, 3, 1\} \Rightarrow \tau\boldsymbol{z}^{(i)} = (2, 3, 1, 1, 2, 3, 2, 1, 1, 3)$ and $S(\tau\boldsymbol{z}^{(i)}, \hat{\boldsymbol{z}}^{\mathrm{MAP}}) = 8$.
  5. $\tau = \{3, 1, 2\} \Rightarrow \tau\boldsymbol{z}^{(i)} = (3, 1, 2, 2, 3, 1, 3, 2, 2, 1)$ and $S(\tau\boldsymbol{z}^{(i)}, \hat{\boldsymbol{z}}^{\mathrm{MAP}}) = 1$.
  6. $\tau = \{3, 2, 1\} \Rightarrow \tau\boldsymbol{z}^{(i)} = (3, 2, 1, 1, 3, 2, 3, 1, 1, 2)$ and $S(\tau\boldsymbol{z}^{(i)}, \hat{\boldsymbol{z}}^{\mathrm{MAP}}) = 3$.

Example: Reordering a single iteration

- Let $n = 10$, $k = 3$, $\boldsymbol{z}^{(i)} = (1,2,3,3,1,2,1,3,3,2)$, $\boldsymbol{\pi}^{(i)} = (\pi_1^{(i)}, \pi_2^{(i)}, \pi_3^{(i)})$ and $\boldsymbol{\theta}^{(i)} = (\theta_1^{(i)}, \theta_2^{(i)}, \theta_3^{(i)})$.
- Suppose that $\hat{\boldsymbol{z}}^{\mathrm{MAP}} = (3,3,1,1,2,3,2,1,1,2)$.
- Find the optimal permutation:
  1. $\tau = \{1,2,3\} \Rightarrow \tau\boldsymbol{z}^{(i)} = (1,2,3,3,1,2,1,3,3,2)$ and $S(\tau\boldsymbol{z}^{(i)}, \hat{\boldsymbol{z}}^{\mathrm{MAP}}) = 1$.
  2. $\tau = \{1,3,2\} \Rightarrow \tau\boldsymbol{z}^{(i)} = (1,3,2,2,1,3,1,2,2,3)$ and $S(\tau\boldsymbol{z}^{(i)}, \hat{\boldsymbol{z}}^{\mathrm{MAP}}) = 2$.
  3. $\tau = \{2,1,3\} \Rightarrow \tau\boldsymbol{z}^{(i)} = (2,1,3,3,2,1,2,3,3,1)$ and $S(\tau\boldsymbol{z}^{(i)}, \hat{\boldsymbol{z}}^{\mathrm{MAP}}) = 2$.
  4. $\tau = \{2,3,1\} \Rightarrow \tau\boldsymbol{z}^{(i)} = (2,3,1,1,2,3,2,1,1,3)$ and $S(\tau\boldsymbol{z}^{(i)}, \hat{\boldsymbol{z}}^{\mathrm{MAP}}) = 8$.
  5. $\tau = \{3,1,2\} \Rightarrow \tau\boldsymbol{z}^{(i)} = (3,1,2,2,3,1,3,2,2,1)$ and $S(\tau\boldsymbol{z}^{(i)}, \hat{\boldsymbol{z}}^{\mathrm{MAP}}) = 1$.
  6. $\tau = \{3,2,1\} \Rightarrow \tau\boldsymbol{z}^{(i)} = (3,2,1,1,3,2,3,1,1,2)$ and $S(\tau\boldsymbol{z}^{(i)}, \hat{\boldsymbol{z}}^{\mathrm{MAP}}) = 3$.
- $\tau = \{2,3,1\} \Rightarrow \tau^{-1} = \{3,1,2\}$, hence $(\boldsymbol{\pi}', \boldsymbol{\theta}')^{(i)} = \big((\pi_3^{(i)}, \pi_1^{(i)}, \pi_2^{(i)}), (\theta_3^{(i)}, \theta_1^{(i)}, \theta_2^{(i)})\big)$.

Example: Reordering a single iteration

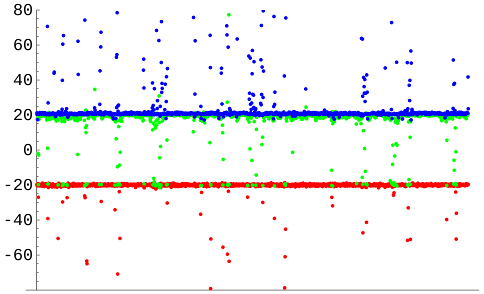
- Let $n = 10$, $k = 3$, $\boldsymbol{z}^{(i)} = (1,2,3,3,1,2,1,3,3,2)$, $\boldsymbol{\pi}^{(i)} = (\pi_1^{(i)}, \pi_2^{(i)}, \pi_3^{(i)})$ and $\boldsymbol{\theta}^{(i)} = (\theta_1^{(i)}, \theta_2^{(i)}, \theta_3^{(i)})$.
- Suppose that $\hat{\boldsymbol{z}}^{\mathrm{MAP}} = (3,3,1,1,2,3,2,1,1,2)$.
- Find the optimal permutation:
  1. $\tau = \{1,2,3\} \Rightarrow \tau\boldsymbol{z}^{(i)} = (1,2,3,3,1,2,1,3,3,2)$ and $S(\tau\boldsymbol{z}^{(i)}, \hat{\boldsymbol{z}}^{\mathrm{MAP}}) = 1$.
  2. $\tau = \{1,3,2\} \Rightarrow \tau\boldsymbol{z}^{(i)} = (1,3,2,2,1,3,1,2,2,3)$ and $S(\tau\boldsymbol{z}^{(i)}, \hat{\boldsymbol{z}}^{\mathrm{MAP}}) = 2$.
  3. $\tau = \{2,1,3\} \Rightarrow \tau\boldsymbol{z}^{(i)} = (2,1,3,3,2,1,2,3,3,1)$ and $S(\tau\boldsymbol{z}^{(i)}, \hat{\boldsymbol{z}}^{\mathrm{MAP}}) = 2$.
  4. $\tau = \{2,3,1\} \Rightarrow \tau\boldsymbol{z}^{(i)} = (2,3,1,1,2,3,2,1,1,3)$ and $S(\tau\boldsymbol{z}^{(i)}, \hat{\boldsymbol{z}}^{\mathrm{MAP}}) = 8$.
  5. $\tau = \{3,1,2\} \Rightarrow \tau\boldsymbol{z}^{(i)} = (3,1,2,2,3,1,3,2,2,1)$ and $S(\tau\boldsymbol{z}^{(i)}, \hat{\boldsymbol{z}}^{\mathrm{MAP}}) = 1$.
  6. $\tau = \{3,2,1\} \Rightarrow \tau\boldsymbol{z}^{(i)} = (3,2,1,1,3,2,3,1,1,2)$ and $S(\tau\boldsymbol{z}^{(i)}, \hat{\boldsymbol{z}}^{\mathrm{MAP}}) = 3$.
- $\tau = \{2,3,1\} \Rightarrow \tau^{-1} = \{3,1,2\}$, hence $(\boldsymbol{\pi}', \boldsymbol{\theta}')^{(i)} = \big((\pi_3^{(i)}, \pi_1^{(i)}, \pi_2^{(i)}), (\theta_3^{(i)}, \theta_1^{(i)}, \theta_2^{(i)})\big)$.
- In order to save computational cost, the optimal permutation can be find applying efficient integer programming algorithms (the assignment problem).

Examples: ECR vs simple methods

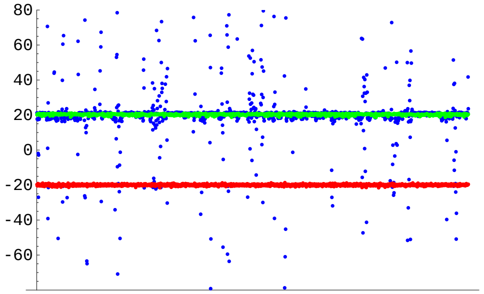
- $n = 160$ simulated observations from $0.10\mathcal{N}(-20, 1) + 0.65\mathcal{N}(20, 3) + 0.25\mathcal{N}(21, 0.5)$

Examples: ECR vs simple methods

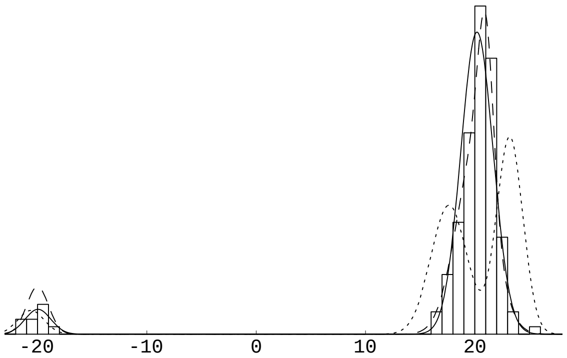
- $n = 160$ simulated observations from
  $0.10\mathcal{N}(-20,1) + 0.65\mathcal{N}(20,3) + 0.25\mathcal{N}(21,0.5)$



PR (Marin, et al., 2005)



ECR



Plug-in density estimate

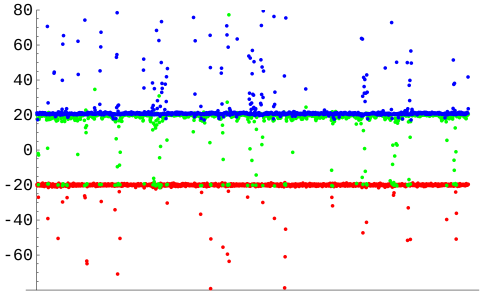
- - - true pdf

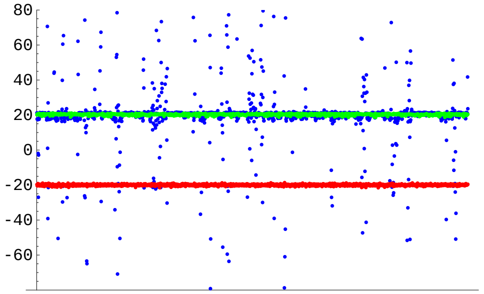
· · · PR

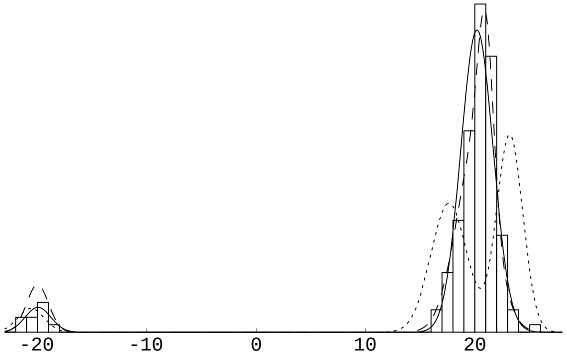
— ECR

Examples: ECR vs simple methods

- $n = 160$ simulated observations from $0.10\mathcal{N}(-20,1) + 0.65\mathcal{N}(20,3) + 0.25\mathcal{N}(21,0.5)$

- 





Plug-in density estimate

- - - true pdf

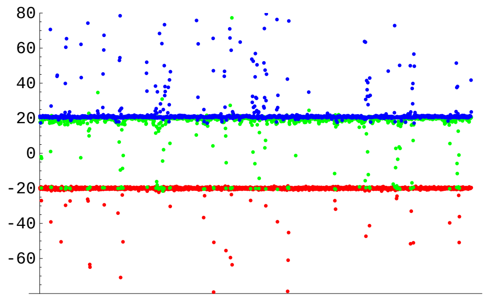
· · · PR

— ECR

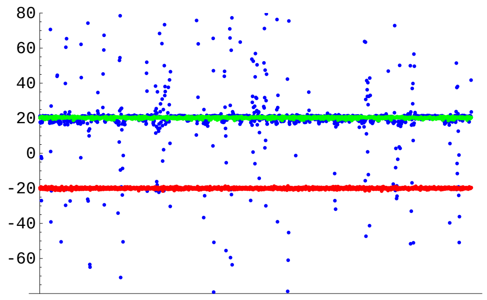
	PR (Marin, et al., 2005)	ECR
$\widehat{\mathbf{E}}(\boldsymbol{\mu}\|\boldsymbol{y})$	$(23.08, 17.62, -20.59)$	$(20.46, 19.61, -19.96)$

Examples: ECR vs simple methods

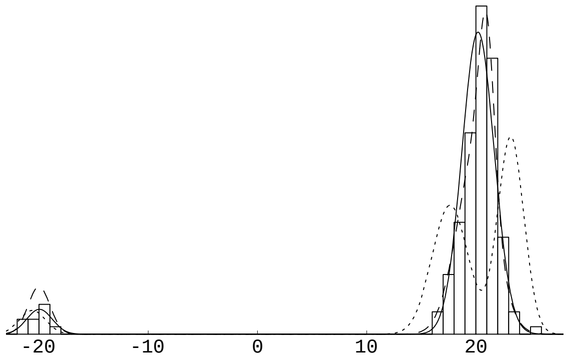
- $n = 160$ simulated observations from
  $0.10\mathcal{N}(-20, 1) + 0.65\mathcal{N}(20, 3) + 0.25\mathcal{N}(21, 0.5)$

- 

  PR (Marin, et al., 2005)

  ECR

  Plug-in density estimate

  - - - true pdf

  · · · PR

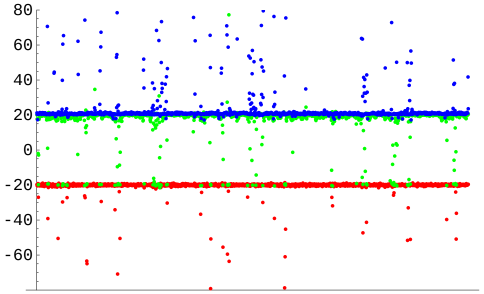
  — ECR

	PR	ECR
$\widehat{\mathbf{E}}(\boldsymbol{\mu}\vert\boldsymbol{y})$	$(23.08, 17.62, -20.59)$	$(20.46, 19.61, -19.96)$
$\widehat{\mathbf{E}}(\boldsymbol{\sigma}^2\vert\boldsymbol{y})$	$(1.83, 2.27, 1.42)$	$(2.01, 2.15, 1.36)$

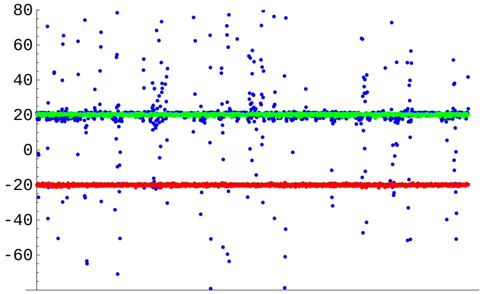
Examples: ECR vs simple methods

- $n = 160$ simulated observations from $0.10\mathcal{N}(-20,1) + 0.65\mathcal{N}(20,3) + 0.25\mathcal{N}(21,0.5)$

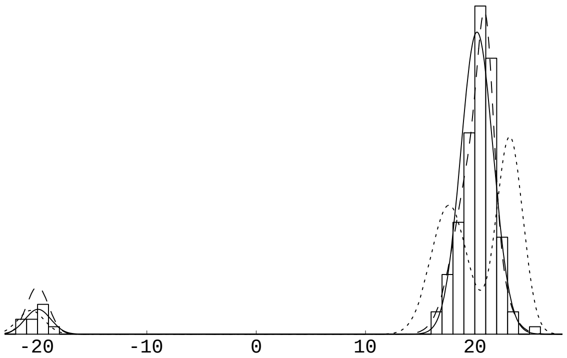
-



PR (Marin, et al., 2005)



ECR



Plug-in density estimate

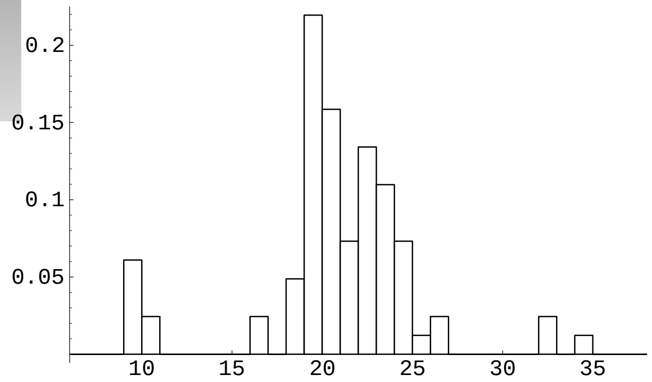
- - - true pdf

· · · PR

— ECR

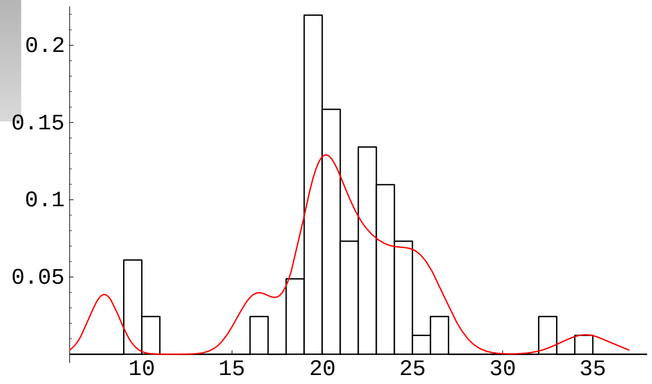
$\widehat{\mathbf{E}}(\boldsymbol{\mu}\vert\boldsymbol{y})$	$(23.08, 17.62, -20.59)$	$(20.46, 19.61, -19.96)$
$\widehat{\mathbf{E}}(\boldsymbol{\sigma}^2\vert\boldsymbol{y})$	$(1.83, 2.27, 1.42)$	$(2.01, 2.15, 1.36)$
$\widehat{\mathbf{E}}(\boldsymbol{\pi}\vert\boldsymbol{y})$	$(.522, .418, .060)$	$(.723, .216, .061)$

Univariate normal mixtures: Galaxy dataset



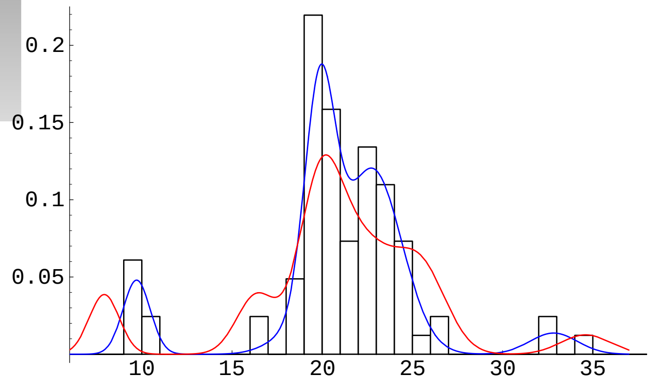
- $n = 82$, $k = 6$, $m = 50000$.

Univariate normal mixtures: Galaxy dataset



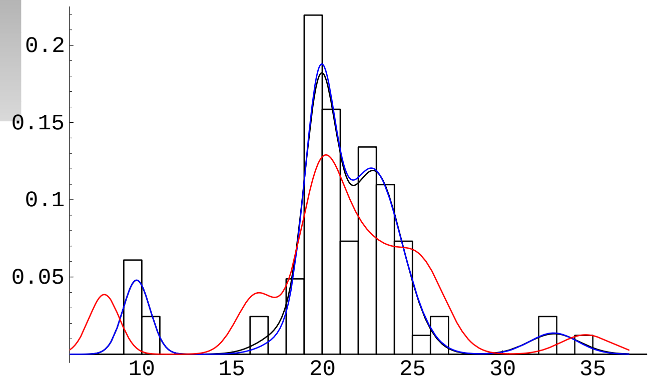
- $n = 82$, $k = 6$, $m = 50000$.
- Pivotal Reordering

Univariate normal mixtures: Galaxy dataset



- $n = 82$, $k = 6$, $m = 50000$.
- Pivotal Reordering
- Kullback - Leibler (256.68 sec)

Univariate normal mixtures: Galaxy dataset



- $n = 82$, $k = 6$, $m = 50000$.
- Pivotal Reordering
- Kullback - Leibler (256.68 sec)
- ECR (1.56 sec)

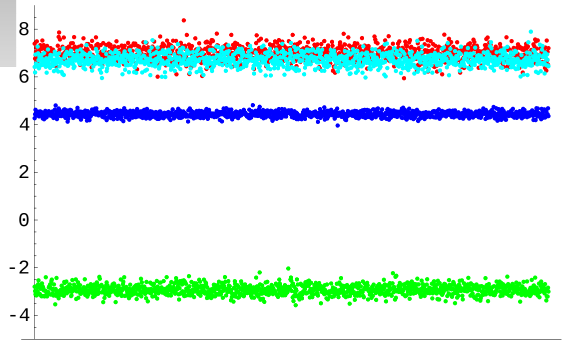
Examples: ECR vs sophisticated methods

$n = 200$ simulated observations from a $k = 4$ component mixture of bivariate normals, $m = 10000$ iterations of the multivariate random beta model (Papageorgiou and Dellaportas, 2005).

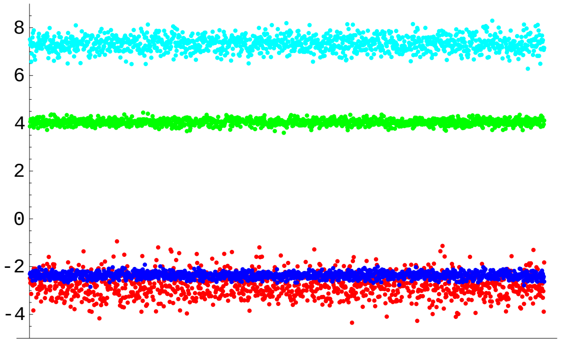
Examples: ECR vs sophisticated methods

parameter	real value	KL (Stephens, 1997)	ECR
$\boldsymbol{\pi}$	$(.25, .25, .25, .25)$	$(.30, .21, .24, .25)$	$(.30, .21, .24, .25)$
$\boldsymbol{\mu}$	$(4.5, -2.5)$	$(4.43, -2.36)$	$(4.43, -2.36)$
	$(-3.0, 4.0)$	$(-2.91, 4.04)$	$(-2.91, 4.04)$
	$(6.5, 7.0)$	$(6.73, 7.34)$	$(6.73, 7.34)$
	$(7.0, -3.0)$	$(6.99, -2.77)$	$(6.99, -2.77)$
$\boldsymbol{\Sigma}$	$\begin{pmatrix} 0.5 & -0.25 \\ -0.25 & 0.5 \end{pmatrix}$	$\begin{pmatrix} 0.54 & -0.20 \\ -0.20 & 0.81 \end{pmatrix}$	$\begin{pmatrix} 0.54 & -0.20 \\ -0.20 & 0.81 \end{pmatrix}$
	$\begin{pmatrix} 0.5 & -0.25 \\ -0.25 & 0.5 \end{pmatrix}$	$\begin{pmatrix} 1.74 & -0.77 \\ -0.77 & 0.69 \end{pmatrix}$	$\begin{pmatrix} 1.74 & -0.77 \\ -0.77 & 0.69 \end{pmatrix}$
	$\begin{pmatrix} 4 & 2.5 \\ 2.5 & 4 \end{pmatrix}$	$\begin{pmatrix} 3.30 & 2.14 \\ 2.14 & 4.09 \end{pmatrix}$	$\begin{pmatrix} 3.30 & 2.14 \\ 2.14 & 4.09 \end{pmatrix}$
	$\begin{pmatrix} 4 & 2.5 \\ 2.5 & 9 \end{pmatrix}$	$\begin{pmatrix} 3.55 & 2.27 \\ 2.27 & 10.13 \end{pmatrix}$	$\begin{pmatrix} 3.55 & 2.27 \\ 2.27 & 10.13 \end{pmatrix}$
CPU Time		16.31	0.45

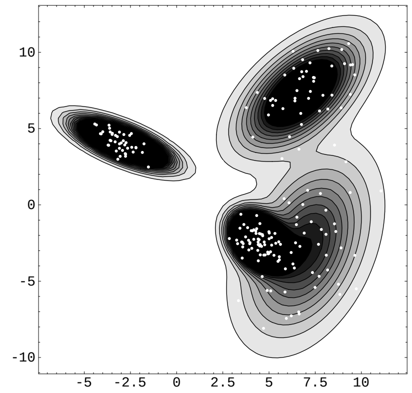
Examples: ECR vs sophisticated methods



μ_{1j}



μ_{2j}



plug-in density estimate

Hidden Markov Model (HMM)

- $Z_1, \ldots, Z_n$: k-state stationary Markov chain with transition matrix $\boldsymbol{P} = (p_{ij})$, $i, j = 1, \ldots, k$, where: $P(z_t = j | z_{t-1} = i) := p_{ij}$.

Hidden Markov Model (HMM)

- $Z_1, \ldots, Z_n$: k-state stationary Markov chain with transition matrix $\boldsymbol{P} = (p_{ij})$, $i, j = 1, \ldots, k$, where: $P(z_t = j | z_{t-1} = i) := p_{ij}$.
- Differences:

Independence Mixture:
For $i = 1, \ldots, n$ and $j = 1, \ldots, k$:

$$\begin{aligned} Y_i | Z_i = j, \boldsymbol{\theta} &\sim f(y|\theta_j) \\ P(Z_i = j | \boldsymbol{\pi}) &= \pi_j \end{aligned}$$

Hidden Markov Model:
For $t = 2, \ldots, n$ and $j = 1, \ldots, k$:

$$\begin{aligned} Y_t | Z_t = j, \boldsymbol{\theta} &\sim f(y|\theta_j) \\ P(Z_t = j | Z_{t-1} = i, \boldsymbol{P}) &= p_{ij} \end{aligned}$$

Hidden Markov Model (HMM)

- $Z_1, \ldots, Z_n$: k-state stationary Markov chain with transition matrix $\boldsymbol{P} = (p_{ij})$, $i, j = 1, \ldots, k$, where: $P(z_t = j | z_{t-1} = i) := p_{ij}$.
- Differences:

Independence Mixture:
For $i = 1, \ldots, n$ and $j = 1, \ldots, k$:

$$\begin{aligned} Y_i | Z_i = j, \boldsymbol{\theta} \quad &\sim \quad f(y|\theta_j) \\ P(Z_i = j | \boldsymbol{\pi}) \quad &= \quad \pi_j \end{aligned}$$

Hidden Markov Model:
For $t = 2, \ldots, n$ and $j = 1, \ldots, k$:

$$\begin{aligned} Y_t | Z_t = j, \boldsymbol{\theta} \quad &\sim \quad f(y|\theta_j) \\ P(Z_t = j | Z_{t-1} = i, \boldsymbol{P}) \quad &= \quad p_{ij} \end{aligned}$$

- Similarity:

Independence Mixture:

$$f(\boldsymbol{\pi}, \boldsymbol{\theta} | \boldsymbol{y}) = \frac{1}{k!} \sum_{\tau \in \mathcal{T}_k} f_{\mathcal{Z}_0} \left(\tau(\boldsymbol{\pi}, \boldsymbol{\theta}) | \boldsymbol{y} \right)$$

Hidden Markov Model:

$$f(\boldsymbol{P}, \boldsymbol{\theta} | \boldsymbol{y}) = \frac{1}{k!} \sum_{\tau \in \mathcal{T}_k} f_{\mathcal{Z}_0} \left(\tau(\boldsymbol{P}, \boldsymbol{\theta}) | \boldsymbol{y} \right)$$

Hidden Markov Model (HMM)

- $Z_1, \ldots, Z_n$: k-state stationary Markov chain with transition matrix $\boldsymbol{P} = (p_{ij})$, $i, j = 1, \ldots, k$, where: $P(z_t = j | z_{t-1} = i) := p_{ij}$.
- Differences:

Independence Mixture:
For $i = 1, \ldots, n$ and $j = 1, \ldots, k$:

$$\begin{aligned} Y_i | Z_i = j, \boldsymbol{\theta} &\sim f(y|\theta_j) \\ P(Z_i = j | \boldsymbol{\pi}) &= \pi_j \end{aligned}$$

Hidden Markov Model:
For $t = 2, \ldots, n$ and $j = 1, \ldots, k$:

$$\begin{aligned} Y_t | Z_t = j, \boldsymbol{\theta} &\sim f(y|\theta_j) \\ P(Z_t = j | Z_{t-1} = i, \boldsymbol{P}) &= p_{ij} \end{aligned}$$

- Similarity:

Independence Mixture:

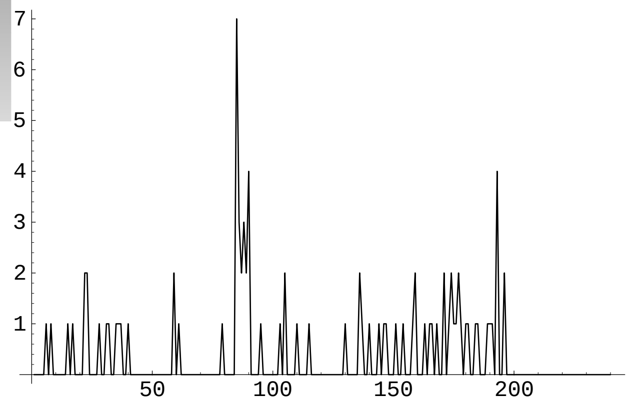
$$f(\boldsymbol{\pi}, \boldsymbol{\theta} | \boldsymbol{y}) = \frac{1}{k!} \sum_{\tau \in \mathcal{T}_k} f_{\mathcal{Z}_0}\left(\tau(\boldsymbol{\pi}, \boldsymbol{\theta}) | \boldsymbol{y}\right)$$

Hidden Markov Model:

$$f(\boldsymbol{P}, \boldsymbol{\theta} | \boldsymbol{y}) = \frac{1}{k!} \sum_{\tau \in \mathcal{T}_k} f_{\mathcal{Z}_0}\left(\tau(\boldsymbol{P}, \boldsymbol{\theta}) | \boldsymbol{y}\right)$$

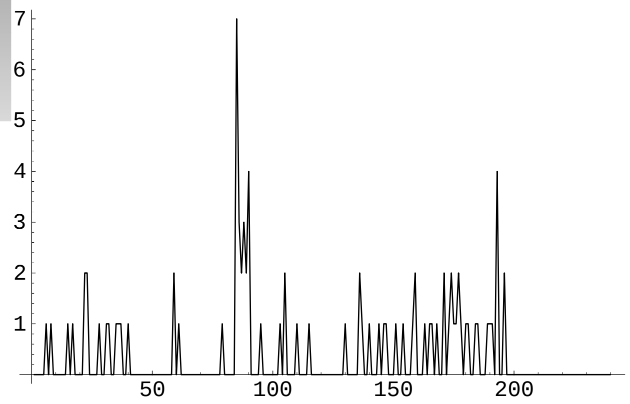
Corollary: ECR algorithm can be applied both to classic mixture models and HMM's as well!

Example: Discrete time inhomogeneus Poisson process



Data: observations of fetal movements in 240 consecutive five second intervals

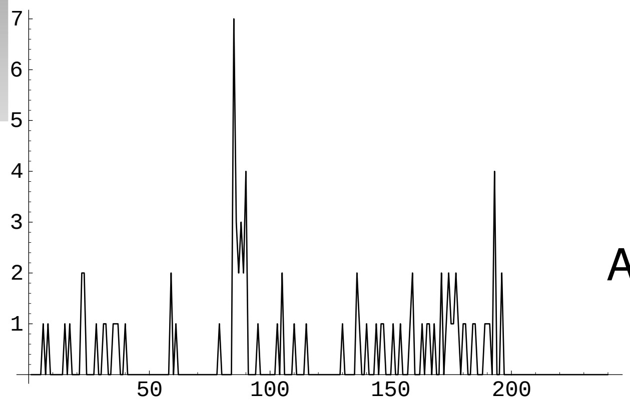
Example: Discrete time inhomogeneus Poisson process



Data: observations of fetal movements in 240 consecutive five second intervals

Model: $Y_t | Z_t = j \sim \mathcal{P}(\lambda_j)$, $t = 1, \ldots, 240$ with $P(Z_t = j | Z_{t-1} = i) = p_{ij}$, $i, j = 1, 2, 3$.

Example: Discrete time inhomogeneus Poisson process

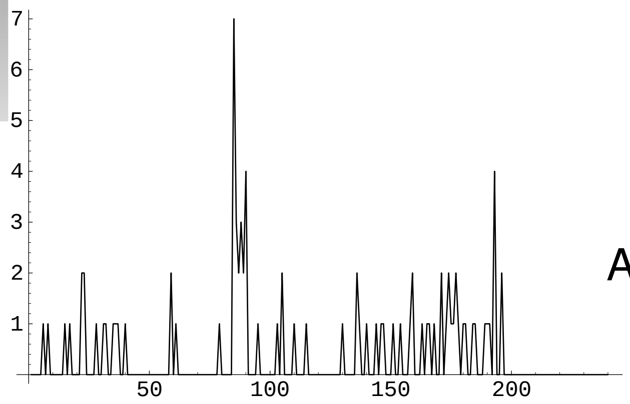


Data: observations of fetal movements in 240 consecutive five second intervals

Model: $Y_t|Z_t = j \sim \mathcal{P}(\lambda_j)$, $t = 1, \dots, 240$ with $P(Z_t = j|Z_{t-1} = i) = p_{ij}$, $i, j = 1, 2, 3$.

A priori: $\lambda_j \sim \mathcal{G}(1, 0.5)$ and $p_{j\cdot} \sim \mathcal{D}(1, 1, 1)$, independently for $j = 1, 2, 3$.

Example: Discrete time inhomogeneus Poisson process



Data: observations of fetal movements in 240 consecutive five second intervals

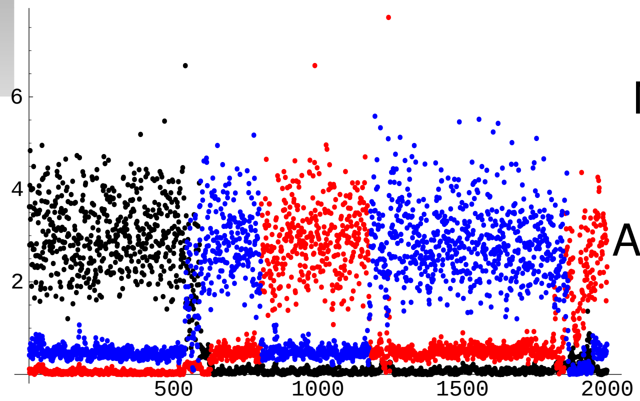
Model: $Y_t|Z_t = j \sim \mathcal{P}(\lambda_j)$, $t = 1, \ldots, 240$ with $P(Z_t = j|Z_{t-1} = i) = p_{ij}$, $i, j = 1, 2, 3$.

A priori: $\lambda_j \sim \mathcal{G}(1, 0.5)$ and $p_{j\cdot} \sim \mathcal{D}(1, 1, 1)$, independently for $j = 1, 2, 3$.

Aim: $f(\lambda_j|\boldsymbol{y})$, and $f(p_{ij}|\boldsymbol{y})$, for $j = 1, \ldots, 3$.

Example: Discrete time inhomogeneus Poisson process

Raw output of $\lambda_1, \lambda_2, \lambda_3$



Data: observations of fetal movements in 240 consecutive five second intervals

Model: $Y_t|Z_t = j \sim \mathcal{P}(\lambda_j)$, $t = 1, \ldots, 240$ with $P(Z_t = j|Z_{t-1} = i) = p_{ij}$, $i, j = 1, 2, 3$.

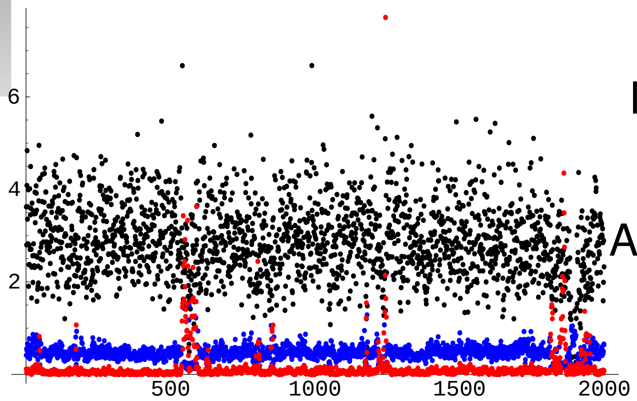
A priori: $\lambda_j \sim \mathcal{G}(1, 0.5)$ and $p_{j\cdot} \sim \mathcal{D}(1, 1, 1)$, independently for $j = 1, 2, 3$.

Aim: $f(\lambda_j|\boldsymbol{y})$, and $f(p_{ij}|\boldsymbol{y})$, for $j = 1, \ldots, 3$.

Gibbs Sampler: 20000 iterations (Label Switching)

Example: Discrete time inhomogeneus Poisson process

Reordered output of $\lambda_1, \lambda_2, \lambda_3$



Data: observations of fetal movements in 240 consecutive five second intervals

Model: $Y_t | Z_t = j \sim \mathcal{P}(\lambda_j)$, $t = 1, \ldots, 240$ with $P(Z_t = j | Z_{t-1} = i) = p_{ij}$, $i, j = 1, 2, 3$.

A priori: $\lambda_j \sim \mathcal{G}(1, 0.5)$ and $p_{j\cdot} \sim \mathcal{D}(1, 1, 1)$, independently for $j = 1, 2, 3$.

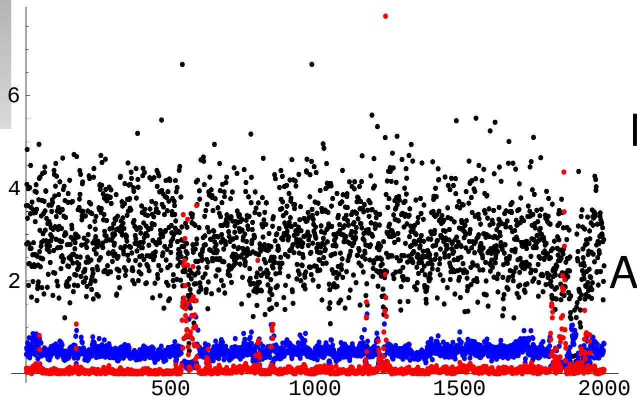
Aim: $f(\lambda_j | \boldsymbol{y})$, and $f(p_{ij} | \boldsymbol{y})$, for $j = 1, \ldots, 3$.

Gibbs Sampler: 20000 iterations (Label Switching)

Reordering: ECR algorithm

Example: Discrete time inhomogeneus Poisson process

Reordered output of $\lambda_1, \lambda_2, \lambda_3$



Data: observations of fetal movements in 240 consecutive five second intervals

Model: $Y_t|Z_t = j \sim \mathcal{P}(\lambda_j)$, $t = 1, \ldots, 240$ with $P(Z_t = j|Z_{t-1} = i) = p_{ij}$, $i, j = 1, 2, 3$.

A priori: $\lambda_j \sim \mathcal{G}(1, 0.5)$ and $p_{j\cdot} \sim \mathcal{D}(1, 1, 1)$, independently for $j = 1, 2, 3$.

Aim: $f(\lambda_j|\boldsymbol{y})$, and $f(p_{ij}|\boldsymbol{y})$, for $j = 1, \ldots, 3$.

Gibbs Sampler: 20000 iterations (Label Switching)

Reordering: ECR algorithm

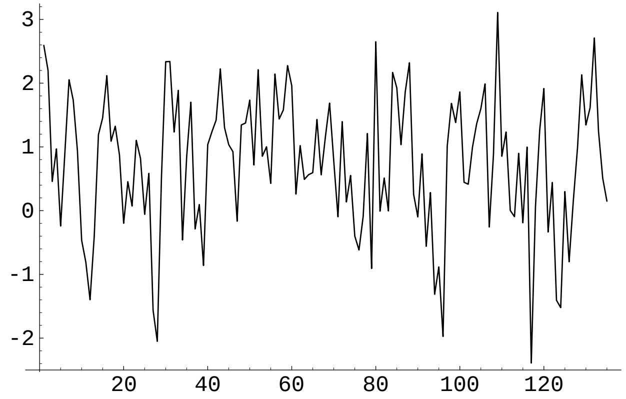
Estimates:

$$\widehat{\mathbb{E}}(\boldsymbol{\lambda}|\boldsymbol{y}) = (0.14, 0.47, 2.85)$$

$$\widehat{\mathbb{E}}(\boldsymbol{P}|\boldsymbol{y}) = \begin{pmatrix} 0.846 & 0.113 & 0.041 \\ 0.090 & 0.883 & 0.027 \\ 0.237 & 0.199 & 0.563 \end{pmatrix}$$

Markovian Switching Autoregressive Model

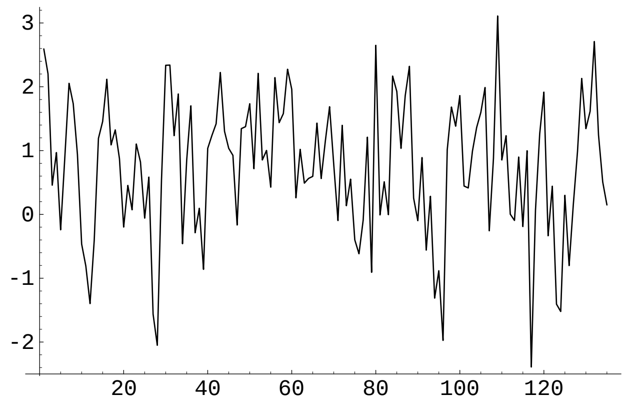
Data: $y_t = 100(\log(\mathrm{GNP}_t) - \log(\mathrm{GNP}_{t-1}))$,
$t = 2, \ldots, 135.$



Markovian Switching Autoregressive Model

Data: $y_t = 100(\log(\mathrm{GNP}_t) - \log(\mathrm{GNP}_{t-1}))$, $t = 2, \ldots, 135$.

Model: $y_t | z_t = j = \beta_{j0} + \sum_{i=1}^{2} \beta_{ji} y_{t-i} + \epsilon_t, t = 1, \ldots, n$, where $\epsilon_t | z_t = j \sim \mathcal{N}(0, \sigma_j^2)$, $j = 1, 2$ ($k = 2$ hidden states, $s = 2$ lagged values, plus constant term).

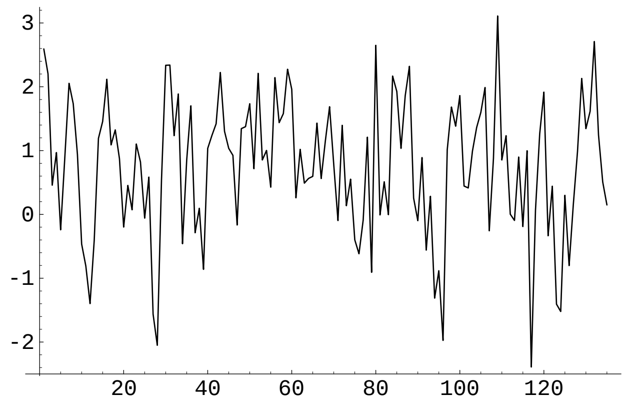


Markovian Switching Autoregressive Model

Data: $y_t = 100(\log(\mathrm{GNP}_t) - \log(\mathrm{GNP}_{t-1}))$, $t = 2, \ldots, 135$.

Model: $y_t|z_t = j = \beta_{j0} + \sum_{i=1}^{2} \beta_{ji} y_{t-i} + \epsilon_t, t = 1, \ldots, n$, where $\epsilon_t|z_t = j \sim \mathcal{N}(0, \sigma_j^2)$, $j = 1, 2$ ($k = 2$ hidden states, $s = 2$ lagged values, plus constant term).

Aim: $f(\beta_{ji}|\boldsymbol{y})$, $f(\sigma_j^2|\boldsymbol{y})$, $j = 1, 2$ and $i = 0, 1, 2$.



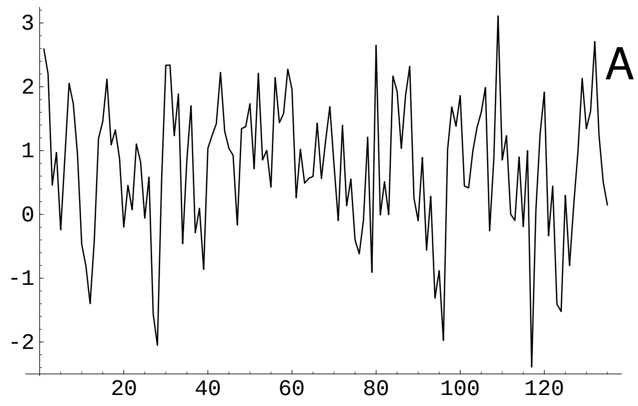
Markovian Switching Autoregressive Model

Data: $y_t = 100(\log(\mathrm{GNP}_t) - \log(\mathrm{GNP}_{t-1}))$, $t = 2, \ldots, 135$.

Model: $y_t | z_t = j = \beta_{j0} + \sum_{i=1}^{2} \beta_{ji} y_{t-i} + \epsilon_t, t = 1, \ldots, n$, where $\epsilon_t | z_t = j \sim \mathcal{N}(0, \sigma_j^2)$, $j = 1, 2$ ($k = 2$ hidden states, $s = 2$ lagged values, plus constant term).

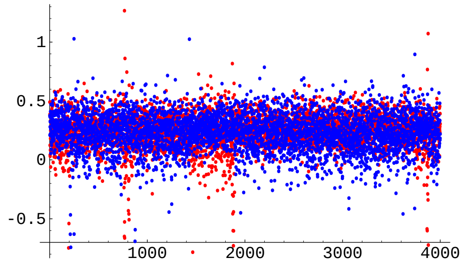
Aim: $f(\beta_{ji} | \boldsymbol{y})$, $f(\sigma_j^2 | \boldsymbol{y})$, $j = 1, 2$ and $i = 0, 1, 2$.

A priori: $\beta_{j0} \sim \mathcal{N}(0, 1)$, $\beta_{j1} \equiv \beta_{j2} \sim \mathcal{N}(0, .25)$, $\sigma_j^{-2} \sim \mathcal{G}(2, 2)$ and $p_{j\cdot} \sim \mathcal{D}(1, 1)$, independently for $j = 1, 2$.

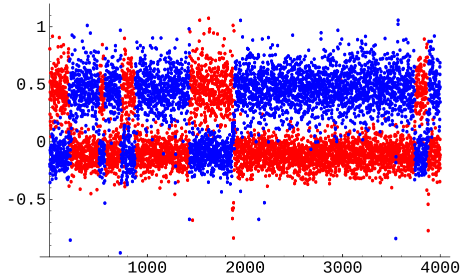


Markovian Switching Autoregressive Model

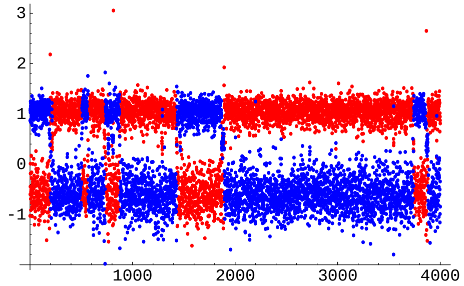
(β_{10}, β_{20})



(β_{11}, β_{21})



(β_{12}, β_{22})



Data: $y_t = 100(\log(\mathrm{GNP}_t) - \log(\mathrm{GNP}_{t-1}))$, $t = 2, \ldots, 135$.

Model: $y_t | z_t = j = \beta_{j0} + \sum_{i=1}^{2} \beta_{ji} y_{t-i} + \epsilon_t, t = 1, \ldots, n$, where $\epsilon_t | z_t = j \sim \mathcal{N}(0, \sigma_j^2)$, $j = 1, 2$ ($k = 2$ hidden states, $s = 2$ lagged values, plus constant term).

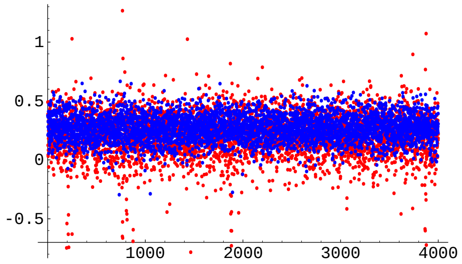
Aim: $f(\beta_{ji} | \boldsymbol{y})$, $f(\sigma_j^2 | \boldsymbol{y})$, $j = 1, 2$ and $i = 0, 1, 2$.

A priori: $\beta_{j0} \sim \mathcal{N}(0, 1)$, $\beta_{j1} \equiv \beta_{j2} \sim \mathcal{N}(0, .25)$, $\sigma_j^{-2} \sim \mathcal{G}(2, 2)$ and $p_{j\cdot} \sim \mathcal{D}(1, 1)$, independently for $j = 1, 2$.

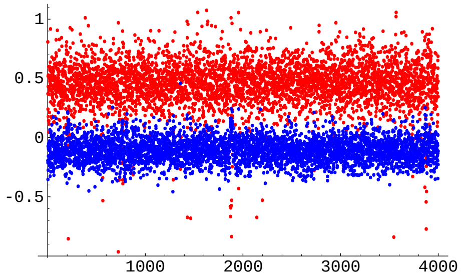
GS: $m = 80000$ (Label Switching)

Markovian Switching Autoregressive Model

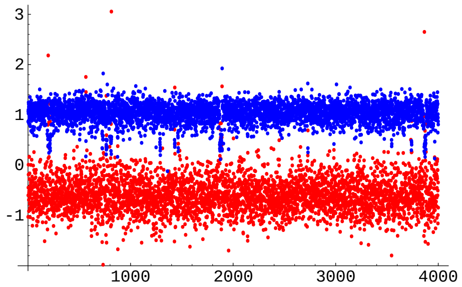
(β_{10}, β_{20})



(β_{11}, β_{21})



(β_{12}, β_{22})



Data: $y_t = 100(\log(\mathrm{GNP}_t) - \log(\mathrm{GNP}_{t-1}))$, $t = 2, \ldots, 135$.

Model: $y_t | z_t = j = \beta_{j0} + \sum_{i=1}^{2} \beta_{ji} y_{t-i} + \epsilon_t, t = 1, \ldots, n$, where $\epsilon_t | z_t = j \sim \mathcal{N}(0, \sigma_j^2)$, $j = 1, 2$ ($k = 2$ hidden states, $s = 2$ lagged values, plus constant term).

Aim: $f(\beta_{ji}|\boldsymbol{y})$, $f(\sigma_j^2|\boldsymbol{y})$, $j = 1, 2$ and $i = 0, 1, 2$.

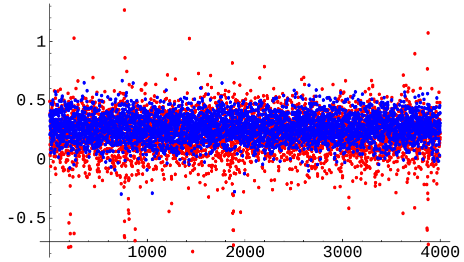
A priori: $\beta_{j0} \sim \mathcal{N}(0, 1)$, $\beta_{j1} \equiv \beta_{j2} \sim \mathcal{N}(0, .25)$, $\sigma_j^{-2} \sim \mathcal{G}(2, 2)$ and $p_{j\cdot} \sim \mathcal{D}(1, 1)$, independently for $j = 1, 2$.

GS: $m = 80000$ (Label Switching)

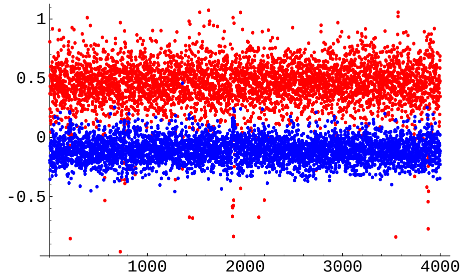
Reordering: ECR algorithm

Markovian Switching Autoregressive Model

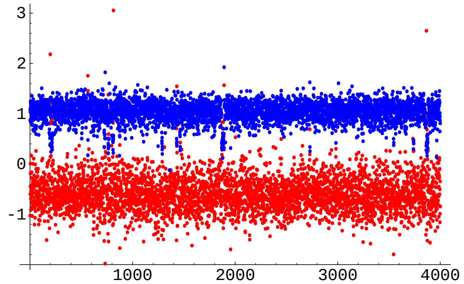
(β_{10}, β_{20})



(β_{11}, β_{21})



(β_{12}, β_{22})



Data: $y_t = 100(\log(\mathrm{GNP}_t) - \log(\mathrm{GNP}_{t-1}))$, $t = 2, \ldots, 135$.

Model: $y_t | z_t = j = \beta_{j0} + \sum_{i=1}^{2} \beta_{ji} y_{t-i} + \epsilon_t, t = 1, \ldots, n$, where $\epsilon_t | z_t = j \sim \mathcal{N}(0, \sigma_j^2)$, $j = 1, 2$ ($k = 2$ hidden states, $s = 2$ lagged values, plus constant term).

Aim: $f(\beta_{ji} | \boldsymbol{y})$, $f(\sigma_j^2 | \boldsymbol{y})$, $j = 1, 2$ and $i = 0, 1, 2$.

A priori: $\beta_{j0} \sim \mathcal{N}(0, 1)$, $\beta_{j1} \equiv \beta_{j2} \sim \mathcal{N}(0, .25)$, $\sigma_j^{-2} \sim \mathcal{G}(2, 2)$ and $p_{j\cdot} \sim \mathcal{D}(1, 1)$, independently for $j = 1, 2$.

GS: $m = 80000$ (Label Switching)

Reordering: ECR algorithm

Estimates: $\widehat{\boldsymbol{\sigma}}^2 = (0.59, 0.53)$, $\quad \widehat{\boldsymbol{P}} = \begin{pmatrix} 0.538 & 0.462 \\ 0.277 & 0.723 \end{pmatrix}$

$$\widehat{\boldsymbol{\beta}} = \begin{pmatrix} 0.21 & 0.46 & -0.57 \\ 0.27 & -0.11 & 1.06 \end{pmatrix}$$

Synopsis

- Marginal Inference $\Leftrightarrow$ Deal with the label switching problem.

Synopsis

- Marginal Inference $\Leftrightarrow$ Deal with the label switching problem.
- Previous work: inefficient or computationally demanding.

Synopsis

- Marginal Inference $\Leftrightarrow$ Deal with the label switching problem.
- Previous work: inefficient or computationally demanding.
- ECR algorithm:

Synopsis

- Marginal Inference $\Leftrightarrow$ Deal with the label switching problem.
- Previous work: inefficient or computationally demanding.
- ECR algorithm:
  1. Theoretical background.

Synopsis

- Marginal Inference $\Leftrightarrow$ Deal with the label switching problem.
- Previous work: inefficient or computationally demanding.
- ECR algorithm:
  1. Theoretical background.
  2. Posterior support remains the same $\Rightarrow$

Synopsis

- Marginal Inference $\Leftrightarrow$ Deal with the label switching problem.
- Previous work: inefficient or computationally demanding.
- ECR algorithm:
  1. Theoretical background.
  2. Posterior support remains the same $\Rightarrow$
  3. ...reveals all the genuine modes of the posterior.

Synopsis

- Marginal Inference $\Leftrightarrow$ Deal with the label switching problem.
- Previous work: inefficient or computationally demanding.
- ECR algorithm:
  1. Theoretical background.
  2. Posterior support remains the same $\Rightarrow$
  3. ...reveals all the genuine modes of the posterior.
  4. Can be applied in any mixture or HM model.

Synopsis

- Marginal Inference $\Leftrightarrow$ Deal with the label switching problem.
- Previous work: inefficient or computationally demanding.
- ECR algorithm:
  1. Theoretical background.
  2. Posterior support remains the same $\Rightarrow$
  3. ...reveals all the genuine modes of the posterior.
  4. Can be applied in any mixture or HM model.
  5. Computational cost independent of the parameter space dimension.

Synopsis

- Marginal Inference $\Leftrightarrow$ Deal with the label switching problem.
- Previous work: inefficient or computationally demanding.
- ECR algorithm:
  1. Theoretical background.
  2. Posterior support remains the same $\Rightarrow$
  3. ...reveals all the genuine modes of the posterior.
  4. Can be applied in any mixture or HM model.
  5. Computational cost independent of the parameter space dimension.
  6. Fast and efficient.

Bibliography

- Celeux, G., Hurn, M. and Robert, C.P. (2000). Computational and inferential difficulties with mixture posterior distributions. *Journal of the American Statistical Association*, **95**, 957–970.
- Diebolt, J. and Robert, C.P. (1994). Estimation of finite mixture distributions through Bayesian sampling. *Journal of the American Statistical Association*, **96**, 194–209.
- Frühwirth-Schnatter, S. (2001). Markov chain Monte Carlo estimation of classical and dynamic switching and mixture models. *Journal of the Royal Statistical Society, Series B*, **56**, 363–375.
- Jasra, A., Holmes, C.C. and Stephens D.A. (2005). Markov Chain Monte Carlo Methods and the Label Switching Problem in Bayesian Mixture Modelling. *Statistical Science*, **20**, 50–67.
- Papastamoulis, P. and Iliopoulos, G. (2009). An artificial allocations based solution to the label switching problem in Bayesian analysis of mixtures of distributions. *Journal Of Computational and Graphical Statistics*, under revision.
- Stephens, M. (2000). Dealing with Label Switching in Mixture Models. *Journal of the Royal Statistical Society Series B*, **62**, 795–809.