Theory and applications of reversible jump MCMC sampling

Aliaksandr Hubin

Trial lecture
University of Oslo

09.11.2018

Outline

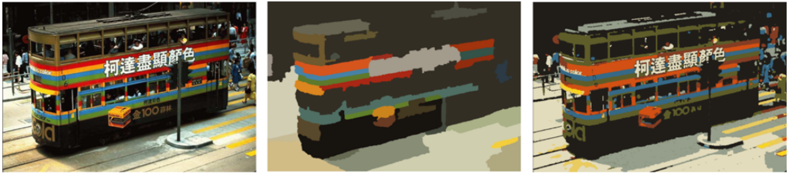
Standard Bayesian inference

Transdimensional Bayesian inference

Simple examples of RJMCMC

Mixtures of Gaussians case study

Discussion



Standard Bayesian inference

Problem

- Considers a parameter $\theta \in \Theta$ of interest;
- Provided likelihood $p(Y|X,\theta) = p(D|\theta)$ and prior $p(\theta)$;
- Aims at computing the posterior $p(\theta|D)$ as:

$$p(\theta|D) = \frac{p(D|\theta)p(\theta)}{\int_{\theta' \in \Theta} p(D|\theta')p(\theta')d\theta'}.$$

Standard Bayesian inference

Problem

- Considers a parameter $\theta \in \Theta$ of interest;
- Provided likelihood $p(Y|X,\theta) = p(D|\theta)$ and prior $p(\theta)$;
- Aims at computing the posterior $p(\theta|D)$ as:

$$p(\theta|D) = \frac{p(D|\theta)p(\theta)}{\int_{\theta' \in \Theta} p(D|\theta')p(\theta')d\theta'}.$$

Methods

- Tractable $p(D) = \int_{\theta' \in \Theta} p(D|\theta')p(\theta')d\theta'$, for example:
  - $Y|X,\theta \sim N(\theta, \sigma^2)$, $\theta \sim N(\mu_0, \sigma_0^2)$;
  - $Y|X,\theta \sim Bernoulli(\theta)$, $\theta \sim Beta(\alpha, \beta)$;
- $p(\theta|D)$ as an ergodic distribution of a Markov Chain:
  - Gibbs sampler;
  - Metropolis-Hastings sampler.

Bayesian inference. Metropolis-Hastings

- A Markov Chain with an ergodic distribution $\pi(\theta) = p(\theta|D)$;
- Transition kernel $p(\theta''|\theta') = q(\theta''|\theta')\alpha(\theta'', \theta')$:
  - $q(\theta''|\theta')$ - proposal probability;
  - $\alpha(\theta'', \theta')$ - acceptance probability:

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

- The detailed balance $\forall A, B \subseteq \Theta$ is satisfied.

$$\blacktriangleright \int_A \int_B \pi(\theta')p(\theta''|\theta')d\theta' d\theta''$$
$$= \int_A \int_B \pi(\theta')q(\theta''|\theta') \min\left(1, \frac{\pi(\theta'')q(\theta'|\theta'')}{\pi(\theta')q(\theta''|\theta')}\right) d\theta' d\theta''$$
$$= \int_B \int_A \pi(\theta'')q(\theta'|\theta'') \min\left(\frac{\pi(\theta')q(\theta''|\theta')}{\pi(\theta'')q(\theta'|\theta'')}, 1\right) d\theta'' d\theta'$$
$$= \int_B \int_A \pi(\theta'')p(\theta'|\theta'')d\theta'' d\theta' \blacktriangleleft$$

Transdimensional Bayesian inference. Problem statement

Problem

- Considers a countable set of parameters $\theta_k \in \Theta_k, k \in \mathcal{K}$;
- Provided $p(D|\theta_k, k)$, $p(\theta_k|k)$ and $p(k)$, $\forall k \in \mathcal{K}$;
- Aims at computing the joint posterior $p(\theta_k, k|D)$ as:

$$p(\theta_k, k|D) = \frac{p(D|\theta_k, k)p(\theta_k|k)p(k)}{\sum_{k' \in \mathcal{K}} p(k') \int_{\theta' \in \Theta} p(D|\theta'_k, k')p(\theta'_k|k')d\theta'}.$$

Transdimensional Bayesian inference. Problem statement

Problem

- Considers a countable set of parameters $\theta_k \in \Theta_k, k \in \mathcal{K}$;
- Provided $p(D|\theta_k, k)$, $p(\theta_k|k)$ and $p(k)$, $\forall k \in \mathcal{K}$;
- Aims at computing the joint posterior $p(\theta_k, k|D)$ as:

$$p(\theta_k, k|D) = \frac{p(D|\theta_k, k)p(\theta_k|k)p(k)}{\sum_{k' \in \mathcal{K}} p(k') \int_{\theta' \in \Theta} p(D|\theta'_k, k')p(\theta'_k|k')d\theta'}.$$

Examples

- Variable dimension models:
  - Model and parameter exploration within some model space $\mathcal{K}$;
  - Inferring on a mixture of Gaussians with feasible mixtures $\mathcal{K}$;
  - *"The number of things you don't know is one of the things you don't know"* [Green and Hastie, 2009]. E.g. tracking unknown number of targets.

Transdimensional Bayesian inference. Model and parameter exploration

1. Consider a class of models $\Omega : m_1(Y|X, \theta_1), ..., m_k(Y|X, \theta_k)$;
2. Put priors for all models $p(m_1), ..., p(m_k)$ and their parameters $p(\theta_1|m_1), ..., p(\theta_k|m_k)$;
3. Obtain the joint posterior distribution of models and parameters $p(m_1, \theta_1|D), ..., p(m_k, \theta_k|D)$;
4. Rao-Blackwellized inference on Δ in the joint space of models and parameters:
   $p(\Delta|D) = \int_\Omega p(m|D) \int_\Theta p(\Delta|m, \theta, D)p(\theta|m, D)d\theta dm$;
5. Easy to extend by means of considering several classes of models $\Omega_1, ..., \Omega_r$ with priors $p(\Omega_1), ..., p(\Omega_r)$.

Transdimensional Bayesian inference. Auxiliary states

- A Markov Chain with an ergodic distribution $p(\theta_k, k|D)$;
- States $\{\theta_k^{(i)} \in \Theta_k, k\}, k \in \mathcal{K}$ might have different domains;
- Use auxiliary states $x \sim g$ and $y \sim f$;
- Construct deterministically a new state $(\theta'', y) = h(\theta', x)$;
- Obtaining a diffeomorphism $(\theta', x) \to (\theta'', y) : \Theta'_x \to \Theta''_y$:
  - Bijection;
  - Differentiable;
  - Dimension matching $\dim \Theta'_x = \dim \Theta''_y$;
- $\left|\frac{\partial(\theta'', y)}{\partial(\theta', x)}\right|$ gives the ratio of the area in Θ''_y space to that in Θ'_x;
- Correct for that in $\alpha((\theta'', k''), (\theta', k'))$ to ensure the detailed balance:

$$\alpha((\theta'', k''), (\theta', k')) = \min\left(1, \frac{\pi(\theta'', k'')f(y)}{\pi(\theta', k')g(x)}\left|\frac{\partial(\theta'', y)}{\partial(\theta', x)}\right|\right).$$

Transdimensional Bayesian inference. Auxiliary states

The detailed balance $\forall A \subseteq \Theta'', B \subseteq \Theta'$ is satisfied:

$$\blacktriangleright \int_{\tilde{A}} \int_{B} \pi(\theta', k') g(x) \min \left(1, \frac{\pi(\theta'', k'') f(y)}{\pi(\theta', k') g(x)} \left| \frac{\partial(\theta'', y)}{\partial(\theta', x)} \right| \right) d\theta' dx$$

$$= \int_{\tilde{A}} \int_{B} \pi(\theta'', k'') f(y) \min \left(\frac{\pi(\theta', k') g(x)}{\pi(\theta'', k'') f(y)} \left| \frac{\partial(\theta', x)}{\partial(\theta'', y)} \right|, 1 \right) \times$$

$$\times \left| \frac{\partial(\theta'', y)}{\partial(\theta', x)} \right| d\theta' dx$$

$$= \int_{\tilde{B}} \int_{A} \pi(\theta'', k'') f(y) \min \left(\frac{\pi(\theta', k') g(x)}{\pi(\theta'', k'') f(y)} \left| \frac{\partial(\theta', x)}{\partial(\theta'', y)} \right|, 1 \right) d\theta'' dy \blacktriangleleft$$

Here by the change of multiple variables in integration:

$$d\theta'' dy = \left| \frac{\partial(\theta'', y)}{\partial(\theta', x)} \right| d\theta' dx$$

Transdimensional Bayesian inference. Mixtures of proposals

- A Markov Chain with an ergodic distribution $p(\theta_k, k|D)$;
- Assume that mixtures of proposals g_i, f_i, with $i \sim Q(\cdot|\theta', k')$;
- Use auxiliary states $x \sim g_i$ and $y \sim f_i$;
- Correct for that in $\alpha((\theta'', k''), (\theta', k'))$ to ensure the detailed balance:

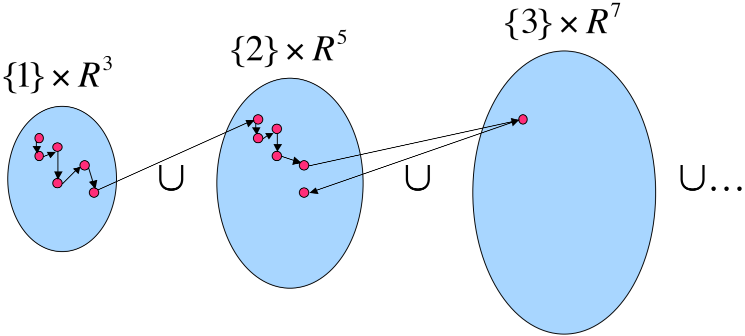
$$\alpha((\theta'', k''), (\theta', k')) = \min\left(1, \frac{\pi(\theta'', k'')Q(i|\theta'', k'')f_i(y)}{\pi(\theta', k')Q(i|\theta', k')g_i(x)}\left|\frac{\partial(\theta'', y)}{\partial(\theta', x)}\right|\right)$$

- Then the detailed balance is satisfied and it can be shown:

$$\int_{\tilde{A}}\int_B \pi(\theta', k')Q(i|\theta'', k'')g_i(x)\alpha_i((\theta'', k''), (\theta', k'))d\theta' dx$$
$$= \int_{\tilde{A}}\int_B \pi(\theta'', k'')Q(i|\theta', k')f_i(y)\alpha_i((\theta', k'), (\theta'', k''))d\theta'' dy.$$

Transdimensional Bayesian inference. Illustration

- A countable set of parameters $\theta_k \in \Theta_k, k \in \mathcal{K}$;
- Each $\Theta_k = \mathcal{R}^{n_k}$;
- The search space $\cup_{k \in \mathcal{K}}(\{k\} \times \mathcal{R}^{n_k})$;
- Illustration from [Dellaert, 2005].



Transdimensional Bayesian inference. Convergence [Brooks and Giudici, 2000]

- Run I chains with different starting points for $2 \cdot T$ iterations;
- Retain θ_i^t for chain $i = \{1, ..., I\}$ at time $t = \{T+1, ..., 2 \cdot T\}$;
- Estimate the total variation V of θ under the target by $\widehat{V}$;
- Estimate the between model variation B_m of θ by $\widehat{B}_m$;
- Estimate the within model variation W_m of θ by $\widehat{W}_m$;
- Estimate the within chain variation W_c of θ by $\widehat{W}_c$;
- $B_m W_c$ gives the within chain variation split between models;
- $W_m W_c$ gives the variance within both chains and models.

Transdimensional Bayesian inference. Convergence [Brooks and Giudici, 2000]

- Run I chains with different starting points for $2 \cdot T$ iterations;
- Retain θ_i^t for chain $i = \{1, ..., I\}$ at time $t = \{T + 1, ..., 2 \cdot T\}$;
- Estimate the total variation V of θ under the target by $\widehat{V}$;
- Estimate the between model variation B_m of θ by $\widehat{B}_m$;
- Estimate the within model variation W_m of θ by $\widehat{W}_m$;
- Estimate the within chain variation W_c of θ by $\widehat{W}_c$;
- $B_m W_c$ gives the within chain variation split between models;
- $W_m W_c$ gives the variance within both chains and models.

1. $\widehat{V}$ and $\widehat{W}_c$ should approximate V well and can be compared;
2. $\widehat{W}_m$ and $\widehat{W}_m \widehat{W}_c$ for within model variance can be compared;
3. $\widehat{B}_m$ and $\widehat{B}_m \widehat{W}_c$ for between model variance can be compared;
4. Detect if running multiple chains and multiple models improve the mixing.

Transdimensional Bayesian inference. Example 1 [Zabaras, 2017]

One dimensional random walk MH

- Consider $\theta_1, \theta_2 \in \mathcal{R}$, $\mathcal{K} = \{1, 2\}$ and $x, y \sim g \in \mathcal{R}$ such that:
  - $\theta_2 = \theta_1 + x, y = -x$;
  - $\theta_1 = \theta_2 + y, x = -y$;
- Then the acceptance ratio becomes:

$$\alpha((\{2\}, \theta_2), (\{1\}, \theta_1)) = \min\left(1, \frac{\pi(\{2\}, \theta_2) g(y)}{\pi(\{1\}, \theta_1) g(x)} \left| \frac{\partial(\theta_2, y)}{\partial(\theta_1, x)} \right| \right)$$
$$= \min\left(1, \frac{\pi(\{2\}, \theta_2) g(\theta_1 - \theta_2)}{\pi(\{1\}, \theta_1) g(\theta_2 - \theta_1)}\right);$$

- Determinant of the Jacobian is equal to 1;
- Considers the same dimensions of θ_1 and θ_2.

Transdimensional Bayesian inference. Example 2 [Zabaras, 2017]

Equilikely birth/death moves within MH

- Consider $\theta_1 \in \mathcal{R}$, $\theta_2 \in \mathcal{R}^2$ and $\mathcal{K} = \{1, 2\}$;
- Consider $x \sim g \in \mathcal{R}$ and $\theta_2 = h(\theta_1, x) = (\theta_1, x)$;
- Then the inverse is $(\theta_1, x) = h^{-1}(\theta_2) = \theta_2$ with probability 1;
- Acceptance of the **birth** move $\{1\} \times \mathcal{R} \to \{2\} \times \mathcal{R}^2$:

$$\alpha((\{2\}, \theta_2), (\{1\}, \theta_1)) = \min\left(1, \frac{\pi(\{2\}, \theta_2) f(y)}{\pi(\{1\}, \theta_1) g(x)} \left|\frac{\partial(\theta_2, y)}{\partial(\theta_1, x)}\right|\right)$$
$$= \min\left(1, \frac{\pi(\{2\}, \theta_2) \times 1}{\pi(\{1\}, \theta_1) g(x)} \left|\frac{\partial\theta_2}{\partial(\theta_1, x)}\right|\right) = \min\left(1, \frac{\pi(\{2\}, \theta_2)}{\pi(\{1\}, \theta_1) g(x)}\right);$$

- Determinant of the Jacobian is equal to 1.

Transdimensional Bayesian inference. Example 2 [Zabaras, 2017]

Equilikely birth/death moves within MH

- Consider $\theta_1 \in \mathcal{R}$, $\theta_2 \in \mathcal{R}^2$ and $\mathcal{K} = \{1, 2\}$;
- Consider $x \sim g \in \mathcal{R}$ and $\theta_2 = h(\theta_1, x) = (\theta_1, x)$;
- Then the inverse is $(\theta_1, x) = h^{-1}(\theta_2) = \theta_2$ with probability 1;
- Acceptance of the **death** move $\{2\} \times \mathcal{R}^2 \to \{1\} \times \mathcal{R}^1$:

$$\alpha((\{1\}, \theta_1), (\{2\}, \theta_2)) = \min\left(1, \frac{\pi(\{1\}, \theta_1) g(x)}{\pi(\{2\}, \theta_2) f(y)} \left| \frac{\partial(\theta_1, x)}{\partial(\theta_2, y)} \right| \right)$$
$$= \min\left(1, \frac{\pi(\{1\}, \theta_1) g(x)}{\pi(\{2\}, \theta_2)} \left| \frac{\partial(\theta_1, x)}{\partial \theta_2} \right| \right) = \min\left(1, \frac{\pi(\{1\}, \theta_1) g(x)}{\pi(\{2\}, \theta_2)} \right);$$

- Here $x = \theta_{2,2}$;
- Determinant of the Jacobian is equal to 1.

Transdimensional Bayesian inference. Example 3 [Zabaras, 2017]

Equilikely split/merge moves within MH

- Consider $\theta_1 \in \mathcal{R}$, $\theta_2 \in \mathcal{R}^2$ and $\mathcal{K} = \{1, 2\}$;
- Consider $x \sim g \in \mathcal{R}$ and $\theta_2 = h(\theta_1, x) = (\theta_1 - x, \theta_1 + x)$;
- Then the inverse is $(\theta_1, x) = h^{-1}(\theta_2) = \left(\frac{\theta_{2,1} - \theta_{2,2}}{2}, \frac{\theta_{2,1} + \theta_{2,2}}{2}\right)$;
- Acceptance of the **split** move $\{1\} \times \mathcal{R} \to \{2\} \times \mathcal{R}^2$:

$$\alpha((\{2\}, \theta_2), (\{1\}, \theta_1)) = \min\left(1, \frac{\pi(\{2\}, \theta_2)}{\pi(\{1\}, \theta_1)g(x)} \times 2\right);$$

Transdimensional Bayesian inference. Example 3 [Zabaras, 2017]

Equilikely split/merge moves within MH

- Consider $\theta_1 \in \mathcal{R}$, $\theta_2 \in \mathcal{R}^2$ and $\mathcal{K} = \{1, 2\}$;
- Consider $x \sim g \in \mathcal{R}$ and $\theta_2 = h(\theta_1, x) = (\theta_1 - x, \theta_1 + x)$;
- Then the inverse is $(\theta_1, x) = h^{-1}(\theta_2) = \left(\frac{\theta_{2,1} - \theta_{2,2}}{2}, \frac{\theta_{2,1} + \theta_{2,2}}{2}\right)$;
- Acceptance of the **split** move $\{1\} \times \mathcal{R} \to \{2\} \times \mathcal{R}^2$:

$$\alpha((\{2\}, \theta_2), (\{1\}, \theta_1)) = \min\left(1, \frac{\pi(\{2\}, \theta_2)}{\pi(\{1\}, \theta_1) g(x)} \times 2\right);$$

- Acceptance of the **merge** move $\{2\} \times \mathcal{R}^2 \to \{1\} \times \mathcal{R}^1$:

$$\alpha((\{1\}, \theta_1), (\{2\}, \theta_2)) = \min\left(1, \frac{\pi(\{1\}, \theta_1) g(x)}{\pi(\{2\}, \theta_2)} \times \frac{1}{2}\right).$$

Image segmentation. Introduction

- We would like to segment different objects on an colored image;
  - E.g. for self-driving vehicles [Romera, 2017];
- Many assume the number of classes for segmentation known;
- This is hardly ever the case in real world;
- Solution: use mixtures of 3 (the number of channels) dimensional Gaussians with unknown number of classes;
- Use RJMCMC for model selection of the number of classes and simultaneous Bayesian inference.

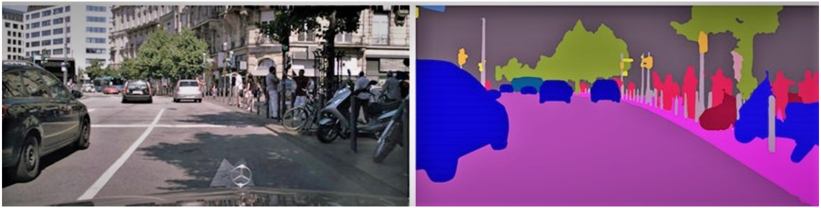


Image segmentation. Mixtures of Gaussians [Kato, 2008]

$$Y_i|\omega_i, \boldsymbol{\mu}_{\omega_i}, \boldsymbol{\Sigma}_{\omega_i}, L, \beta, p_{\omega_i} \overset{\text{ind}}{\sim} N_3(\mu_{\omega_i}, \boldsymbol{\Sigma}_{\omega_i});$$

$$p(\omega_i|\beta, p_{\omega_i}, L) = \frac{1}{Z}\exp\left(-\sum_{j\in\mathcal{P}} V(\omega_i, \omega_j)\right) p_{\omega_i}, p_{\omega_i} = \frac{1}{L};$$

$$V(\omega_i, \omega_j) = \beta \cdot \delta(\omega_i, \omega_j);$$

$$\delta(\omega_i, \omega_j) = (\mathrm{I}(\omega_i \neq \omega_j) - \mathrm{I}(\omega_i = \omega_j));$$

$$L \sim Unif(1, L_{max}), \mu_{\omega_i}|L \sim Unif(\{M\}), \boldsymbol{\Sigma}_{\omega_i}|L \sim Unif(\{S\}).$$

- $\omega_i \in \{1, ..., L\}, i \in \mathcal{P}$ are hidden segmentations for pixels $\mathcal{P}$;
- Here the models are identified by the number of mixtures L;
- Hence we have a model space $\{1, ..., L_{max}\}, L_{max} = 50$;
- $\beta = 2.5$ is the interaction strength parameter;
- $\{M\}$ and $\{S\}$ are the domains for means and covariances;
- p_λ are probability of a pixel be from class λ, $p_1 + ... + p_L = 1$.

Image segmentation. Splitting classes [Kato, 2008]

1. Choose a class λ' to split with a uniform probability $p_{ss} = \frac{1}{L'}$;
2. Make a split $L'' = L' + 1$ and $\lambda'' = (\lambda''_1, \lambda''_2)$, $x \sim g \in \mathcal{R}^{13}_{[0,1]}$;
3. Increase dimensionality in $\theta'' = \{\vec{p''}, \vec{\boldsymbol{\mu}''}, \vec{\boldsymbol{\Sigma}''}\}$ via $h(\theta', x)$ as:
   - $p''_{\lambda''_1} = p'_{\lambda'} u$, $p''_{\lambda''_2} = p'_{\lambda'}(1-u)$;
   - $\mu''_{\lambda''_1,i} = \mu'_{\lambda',i} + w_i \sqrt{\Sigma_{\lambda',i,i} \frac{1-u}{u}}$, $\mu'_{\lambda'_2,i} = \mu'_{\lambda',i} - w_i \sqrt{\Sigma_{\lambda',i,i} \frac{u}{1-u}}$;
   - $\Sigma''_{\lambda''_1,i,j} = \begin{cases} z_{i,i}(1-w_i^2)\Sigma_{\lambda',i,i}\frac{1}{u}, \text{if } i = j; \\ z_{i,j}\Sigma_{\lambda',i,j}\sqrt{(1-w_i^2)(1-w_j^2)z_{i,i}z_{j,j}}, \text{if } i \neq j; \end{cases}$
   - $\Sigma''_{\lambda''_2,i,j} =$
     $\begin{cases} (1-z_{i,i})(1-w_i^2)\Sigma_{\lambda',i,i}\frac{1}{u}, \text{if } i = j; \\ (1-z_{i,j})\Sigma_{\lambda',i,j}\sqrt{(1-w_i^2)(1-w_j^2)(1-z_{i,i})(1-z_{j,j})}, \text{if } i \neq j; \end{cases}$
   - Here $u, w_i, z_{i,j} \sim Unif(0,1)$ and $x = (u, \boldsymbol{w}, \boldsymbol{z})$;
4. Design of the proposals is based on the moment matching conditions allowing to build reasonably good proposals in RJMCMC
   [Richardson and Green, 1997, Fan and Sisson, 2011].

Image segmentation. Merging classes [Kato, 2008]

1. Choose a pair $(\lambda_1'', \lambda_2'')$ to merge w.r.t. $p_{ms} = \frac{1/d(\lambda_1'',\lambda_2'')}{\sum_{\lambda,\kappa\in\Lambda} 1/d(\lambda,\kappa)}$;
   - Here $d(\lambda, \kappa)$ is the Mahalanobis distance between λ and κ;
2. Make a merge $L' = L'' - 1$ and λ', $y \sim f$;
3. Decrease dimensionality in $\theta' = \{\vec{p'}, \vec{\boldsymbol{\mu}'}, \vec{\boldsymbol{\Sigma}'}\}$ via $h^{-1}(\theta'')$ similarly to the split move based on the solution of the moment matching conditions:
   - $p'_{\lambda'} = p''_{\lambda_1''} + p''_{\lambda_2''}$;
   - $p'_{\lambda'}\boldsymbol{\mu}'_{\lambda'} = p''_{\lambda_1''}\boldsymbol{\mu}''_{\lambda_1''} + p''_{\lambda_2''}\boldsymbol{\mu}''_{\lambda_2''}$;
   - $p'_{\lambda'}(\boldsymbol{\mu}'_{\lambda'}\boldsymbol{\mu'}^{\mathsf{T}}_{\lambda'} + \boldsymbol{\Sigma}'_{\lambda'}) =$ $p''_{\lambda_1''}(\boldsymbol{\mu}''_{\lambda_1''}\boldsymbol{\mu''}^{\mathsf{T}}_{\lambda_1''} + \boldsymbol{\Sigma}''_{\lambda_1''}) + p''_{\lambda_2''}(\boldsymbol{\mu}''_{\lambda_2''}\boldsymbol{\mu''}^{\mathsf{T}}_{\lambda_2''} + \boldsymbol{\Sigma}''_{\lambda_2''})$.

Image segmentation. Merging classes [Kato, 2008]

1. Choose a pair $(\lambda_1'', \lambda_2'')$ to merge w.r.t. $p_{ms} = \frac{1/d(\lambda_1'',\lambda_2'')}{\sum_{\lambda,\kappa\in\Lambda} 1/d(\lambda,\kappa)}$;
   - Here $d(\lambda, \kappa)$ is the Mahalanobis distance between λ and κ;
2. Make a merge $L' = L'' - 1$ and λ', $y \sim f$;
3. Decrease dimensionality in $\theta' = \{\vec{p'}, \vec{\boldsymbol{\mu}'}, \vec{\boldsymbol{\Sigma}'}\}$ via $h^{-1}(\theta'')$ similarly to the split move based on the solution of the moment matching conditions:
   - $p'_{\lambda'} = p''_{\lambda_1''} + p''_{\lambda_2''}$;
   - $p'_{\lambda'}\boldsymbol{\mu}'_{\lambda'} = p''_{\lambda_1''}\boldsymbol{\mu}''_{\lambda_1''} + p''_{\lambda_2''}\boldsymbol{\mu}''_{\lambda_2''}$;
   - $p'_{\lambda'}(\boldsymbol{\mu}'_{\lambda'}\boldsymbol{\mu}'^{\mathsf{T}}_{\lambda'} + \boldsymbol{\Sigma}'_{\lambda'}) =$
     $p''_{\lambda_1''}(\boldsymbol{\mu}''_{\lambda_1''}\boldsymbol{\mu}''^{\mathsf{T}}_{\lambda_1''} + \boldsymbol{\Sigma}''_{\lambda_1''}) + p''_{\lambda_2''}(\boldsymbol{\mu}''_{\lambda_2''}\boldsymbol{\mu}''^{\mathsf{T}}_{\lambda_2''} + \boldsymbol{\Sigma}''_{\lambda_2''})$.

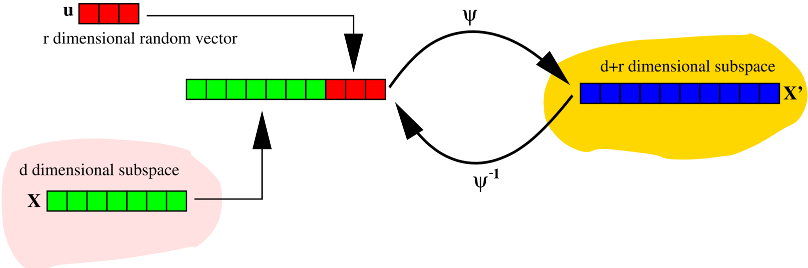


Image segmentation. Acceptance probabilities [Kato, 2008]

- Acceptance probabilities:
  - For split move is $\alpha(L'', \theta'', L', \theta') = \min(1, A)$;
  - For merge move is $\alpha(L', \theta', L'', \theta'') = \min(1, \frac{1}{A})$;
- Here A, based on the standard RJMCMC theory, reduces to:

$$A = \frac{p(L'', \theta'' | \boldsymbol{Y})}{p(L', \theta' | \boldsymbol{Y})} \times \frac{p_m(L'') p_{ms}(\lambda''_1, \lambda''_2)}{p_s(L') p_{ss}(\lambda') p_r} \times \frac{1}{p(u) \prod_{i=1}^{3} p(w_i) \prod_{j=1}^{3} p(z_{i,j})}$$

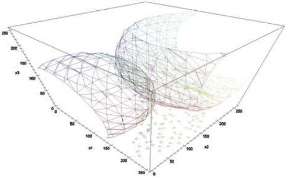
$$\times -p'_{\lambda'} \prod_{i=1}^{3} \left(\frac{\boldsymbol{\Sigma}'_{\lambda',i,i}}{u(1-u)} (1 - w_i^2)(1 - z_{i,i}) z_{i,i} \prod_{j=1}^{3} \frac{\boldsymbol{\Sigma}'_{\lambda',i,j}}{u(1-u)} \right);$$

$$p_r = \prod_{s : \omega''_s \in \{\lambda''_1, \lambda''_2\}} \mathcal{N}_3(Y_s | \mu''_{\omega''_s}, \boldsymbol{\Sigma}''_{\omega''_s}) \times p''_{\omega''_s} \exp(-\beta \sum_{r \in \mathcal{P}} \delta(\omega''_s, \omega''_r)).$$

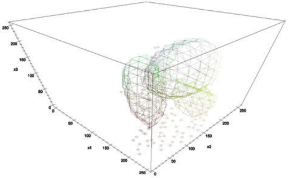
Image segmentation. Illustrative results [Kato, 2008]

Original image

Segmentation result (3 labels)



Initial Gaussians



Final estimation (3 classes)

Image segmentation. What is behind [Zhang et al., 2004]

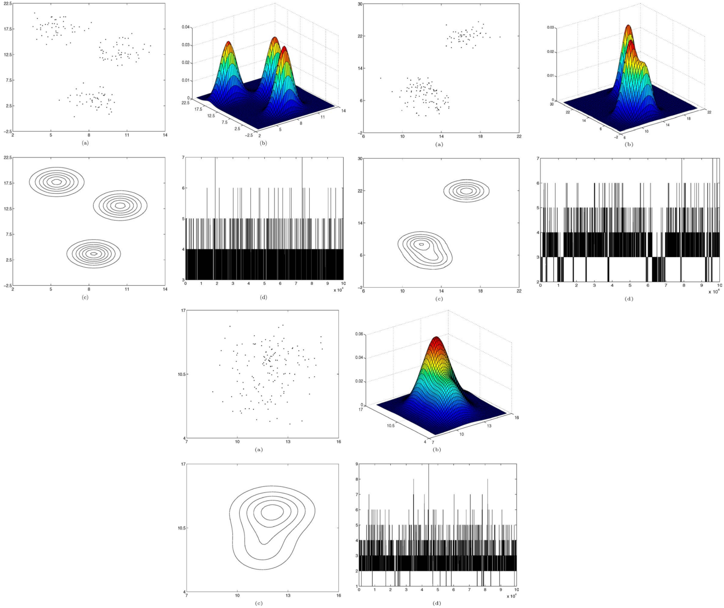
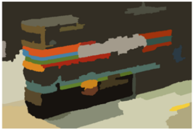


Image segmentation. Further results [Kato, 2008]

Original image



JSEG (37 labels)

RJMCMC (10 labels)

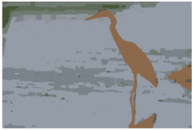
Original image

JSEG (50 labels)

RJMCMC (8 labels)

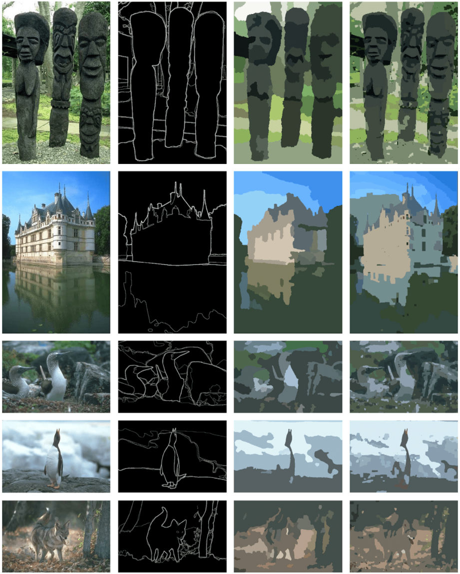
Original image

JSEG (14 labels)



RJMCMC (9 labels)

Image segmentation. Further results [Kato, 2008]



Original image | Human segmentation | JSEG | RJMCMC

RJMCMC. Interesting applications

- Tracking a variable number of interacting targets [Khan et al., 2005];
- Change points detection analysis [Zhao and Chu, 2010];
- Bayesian model determination [Green, 1995];
- Building configurations of Bayesian ANN [Andrieu et al., 2000];
- Inference on hidden Markov models [Robert et al., 2000];
- Synthesizing open worlds with constraints [Yeh et al., 2012].

RJMCMC. Actual problems and open questions

- Designing efficient RJMCMC is challenging;
- Multi-modal space exploration is especially challenging;
- Constructing proposals and calculation of the acceptance probabilities is often tedious in real life applications;
- Efficient subsampling techniques must be developed to be able to work with *"tall data"*;
- Convergence assessment and validity of the output analysis is far from obvious in ultrahigh-dimensional settings.

RJMCMC. Modifications

- Use moments matching proposals [Richardson and Green, 1997, Fan and Sisson, 2011];
- Use a secondary Markov chain to make a proposal [Al-Awadhi et al., 2004];
- Target tempering to enhance mixing properties [Karagiannis and Andrieu, 2013];
- Parallel tempering (population based) chains [Jasra et al., 2007, Fouskakis et al., 2009];
- State space augmentation to the largest model [Brooks et al., 2003];
- General usage of auxiliary states in RJMCMC [Brooks et al., 2003].

References I

Al-Awadhi, F., Hurn, M., and Jennison, C. (2004).
Improving the acceptance rate of reversible jump mcmc proposals.
Statistics & probability letters, 69(2):189–198.

Andrieu, C., De Freitas, N., and Doucet, A. (2000).
Reversible jump mcmc simulated annealing for neural networks.
In *Proceedings of the Sixteenth conference on Uncertainty in artificial intelligence*, pages 11–18. Morgan Kaufmann Publishers Inc.

Brooks, S. P. and Giudici, P. (2000).
Markov chain monte carlo convergence assessment via two-way analysis of variance.
Journal of Computational and Graphical Statistics, 9(2):266–285.

Brooks, S. P., Giudici, P., and Roberts, G. O. (2003).
Efficient construction of reversible jump markov chain monte carlo proposal distributions.
Journal of the Royal Statistical Society: Series B (Statistical Methodology), 65(1):3–39.

Dellaert, F. (2005).
ICCV05 Tutorial: MCMC for vision.

Fan, Y. and Sisson, S. A. (2011).
Reversible jump mcmc.
Handbook of Markov Chain Monte Carlo, pages 67–92.

Fouskakis, D., Ntzoufras, I., and Draper, D. (2009).
Population-based reversible jump markov chain monte carlo methods for bayesian variable selection and evaluation under cost limit restrictions.
Journal of the Royal Statistical Society: Series C (Applied Statistics), 58(3):383–403.

Green, P. J. (1995).
Reversible jump markov chain monte carlo computation and bayesian model determination.
Biometrika, 82(4):711–732.

References II

Green, P. J. and Hastie, D. I. (2009).
Reversible jump MCMC.
Genetics, 155(3):1391–1403.

Jasra, A., Stephens, D. A., and Holmes, C. C. (2007).
Population-based reversible jump markov chain monte carlo.
Biometrika, 94(4):787–807.

Karagiannis, G. and Andrieu, C. (2013).
Annealed importance sampling reversible jump mcmc algorithms.
Journal of Computational and Graphical Statistics, 22(3):623–648.

Kato, Z. (2008).
Segmentation of color images via reversible jump MCMC sampling.
Image and Vision Computing, 26(3):361–371.

Khan, Z., Balch, T., and Dellaert, F. (2005).
Mcmc-based particle filtering for tracking a variable number of interacting targets.
IEEE transactions on pattern analysis and machine intelligence, 27(11):1805–1819.

Richardson, S. and Green, P. J. (1997).
On bayesian analysis of mixtures with an unknown number of components (with discussion).
Journal of the Royal Statistical Society: series B (statistical methodology), 59(4):731–792.

Robert, C. P., Ryden, T., and Titterington, D. M. (2000).
Bayesian inference in hidden markov models through the reversible jump markov chain monte carlo method.
Journal of the Royal Statistical Society: Series B (Statistical Methodology), 62(1):57–75.

Romera, E. (2017).
Pytorch code for semantic segmentation using erfnet.
`github.com/Eromera/erfnet_pytorch`.

References III

Yeh, Y.-T., Yang, L., Watson, M., Goodman, N. D., and Hanrahan, P. (2012).
Synthesizing open worlds with constraints using locally annealed reversible jump mcmc.
ACM Transactions on Graphics (TOG), 31(4):56.

Zabaras, N. (2017).
Variable Dimension Models: Reversible Jump MCMC.

Zhang, Z., Chan, K. L., Wu, Y., and Chen, C. (2004).
Learning a multivariate Gaussian mixture model with the reversible jump MCMC algorithm.
Statistics and Computing, 14(4):343–355.

Zhao, X. and Chu, P.-S. (2010).
Bayesian changepoint analysis for extreme events (typhoons, heavy rainfall, and heat waves): An rjmcmc approach.
Journal of Climate, 23(5):1034–1046.