Reversible Jump Markov Chain Monte Carlo

- Useful in inference problems where the number of parameters change.

- How do we jump between different subspaces of different dimensions while conserving detailed balance.

→ Motivation in terms of Model selection

- $\{M_k, k \in \mathbb{N}\}$ ← countable collection of models.

- Each M_k has a vector $\underset{\text{unknown parameters}}{\theta^{(k)}} \in \mathbb{R}^{n_k}$, where n_k is the size of $\theta^{(k)}$.

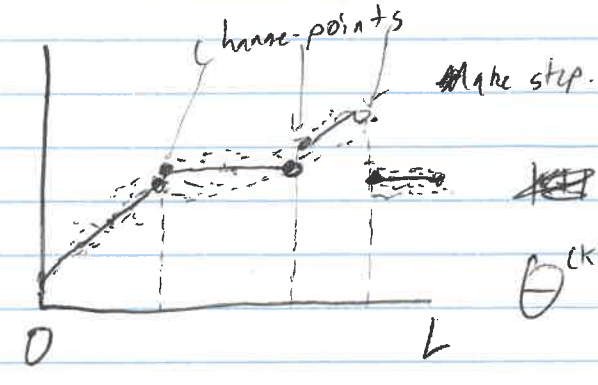
- $(k, \theta^{(k)}, \underset{\text{data}}{y})$ has joint distribution

$$p(k, \theta^{(k)}, y) = \underbrace{p(k)}_{\text{model prob}} \underbrace{p(\theta^{(k)} \mid k)}_{\text{prior}} \underbrace{p(y \mid k, \theta^{(k)})}_{\text{likelihood}}.$$

Let $x = (k, \theta^{(k)})$. For a specific k, $x \in C_k = \{k\} \times \mathbb{R}^{n_k}$

combined parameter space.

x varies over the space $C = \bigcup_{k \in \mathbb{N}} C_k$.



A Model M_k has k change-points.

$\theta^{(k)}$ = k change-point positions and $(k+1)$ values for each subinterval,

$n_k = k + k + 1 = 2k + 1$.

Joint Posterior $p(K, \theta^{(K)} | y) = p(k|y)\, p(\theta^{(k)} | K, y)$

$\uparrow$ model indicator.

$p(k|y) \rightarrow \frac{p(k_1|y)}{p(k_0|y)} \div \frac{p(k_1)}{p(k_0)}$ If you can reduce down to two models.

MCMC

$\pi(dx)$ is the target distribution. $\int_A \pi(dx) = \pi(x \in A)$.

Markov transition kernel ~~needed~~ $P(x, dx')$

1. A periodic
2. Irreducible
3. Detailed Balance $\rightarrow \int_A \int_B \pi(dx) P(x, dx') = \int_B \int_A \pi(dx') P(x', dx)$

$\uparrow$ want to conserve.

for all ~~approp~~ appropriate A, B, Borel sets?

$$\alpha(x, x') = \min\left\{1, \frac{\pi(x_T' | x_{-T})\, q_T(x_T; x')}{\pi(x_T | x_{-T})\, q_T(x_T'; x)}\right\}$$

General

At current x, propose move m, that transitions to state dx'. That has probability $q_m(x, dx')$

$\Sigma_m q_m(x, C) \leq 1$ and there is $1 - \Sigma_m q_m(x, C)$ probability of no change.

For certain m, $q_m(x, c) = 0$.

$\alpha_m(x, x') \leftarrow$ acceptance prob of accepting x' given move m,

Transition Kernel

$$P(x, B) = \sum_m \int_B q_m(x, dx')\, \alpha_m(x, x') + s(x)\, I(x \in B)$$

$$s(x) := \sum_m \int_C q_m(x, dx')\{1 - \alpha_m(x, x')\} + 1 - \sum_m q_m(x, C)$$

$\uparrow$ prob of x not moving.

Borel set

Any set in a space that can ~~be~~ be formed from open sets (or closed sets) through the operations of countable union, countable intersection, relative complement.

The collection of All Borel sets on a space X forms a σ-algebra $\mathcal{B}$. $P: \mathcal{B} \to [0,1]$

$$\int_{A}\int_{B} \pi(dx) P(x, dx') = \int_{B}\int_{A} \pi(dx') P(x', dx)$$

Need:

$$\int_A \pi(dx) \sum_m \int_B q_m(x, dx') \alpha_m(x, x') + s(x) I(x \in B)$$

sufficient

$$\boxed{\sum_m \int_A \pi(dx) \int_B q_m(x, dx') \alpha_m(x, x') + \int_{A \cap B} \pi(dx) s(x) = \sum_m \int_B \pi(dx') \int_A q_m(x', dx) \alpha_m(x', x) + \int_{B \cap A} \pi(dx') s(x')}$$

~~Choose $\alpha_m(x, x')$ s.t.~~

First assume $\pi(dx) q_m(x, dx')$ has a finite density $f_m(x, x')$ with respect to a symmetric measure ξ_m on $C \times C$.

$\Downarrow$

$$\int_A \pi(dx) \int_B q_m(x, dx') \alpha_m(x, x')$$

$$= \int_A \int_B \xi_m(dx, dx') f_m(x, x') \alpha_m(x, x')$$

$$= \int_B \int_A \xi_m(dx', dx) f_m(x', x) \alpha_m(x', x)$$

$$= \int_B \pi(dx') \int_A q_m(x', dx) \alpha_m(x', x)$$

← Dimension Matching assumption.

If $\alpha_m(x, x') f_m(x, x') = \alpha(x', x) f_m(x', x)$

$$\Rightarrow \alpha_m(x, x') = \min\left\{1, \frac{f_m(x', x)}{f_m(x, x')}\right\} = \min\left\{1, \frac{\pi(dx') q_m(x', dx)}{\pi(dx) q_m(x, dx')}\right\}$$

1. The sampling method is entirely constructive, You can exploit the structure of the problem etc.

2. π does not need to be normalised, relative normalising constants between different subspaces are needed.

3. Specifically, priors $p(\theta^{(k)}|k)$ don't need to be properly normalized, but there can only be one unknown multiplicative constant among them, unless only posteriors conditional on k are needed.

4. If only one subspace $\Rightarrow$ just metropolis -hastings.

Ex: $C_1 = \{1\} \times \mathbb{R}$ and $C_2 = \{2\} \times \mathbb{R}^2$, with π having proper density on each subspace given $k=1$ and $k=2$.

A good move might be from $(2, \theta_1, \theta_2) \in C_2$ to $\{1, \frac{1}{2}(\theta_1 + \theta_2)\}$

$$\int_B \pi(dx) \int_A q_m(x, dx') \text{ where } A \subset C_1 \text{ and } B \subset C_2$$

must have density with respect to a measure on $k \times \mathbb{R}^2$ placing all mass on $\{(\theta, \theta_1, \theta_2) : \theta = \frac{1}{2}(\theta_1 + \theta_2)\}$.

To attain detailed balance, the reverse move from A to B should be defined by $q_m(x, dx')$ that is singular for each $x = (1, \theta)$ with all its mass on $\{(2, \theta_1, \theta_2) : \theta = \frac{1}{2}(\theta_1 + \theta_2)\}$.

Draw u from some dist H from θ and set $\theta_1 = \theta + u$, $\theta_2 = \theta - u$, More standard:

Suppose two subspaces, given $K=1$ and $K=2$.

$P(\theta^{(1)}|K=1)$, $P(\theta^{(2)}|K=2)$ are proper densities in R^{n_1} and R^{n_2}, respectively.

Consider just one move type, s.t. $q(X, C_1)=0$, $X \in C_1$ and $q(X, C_2)=0$, $X \in C_2$. You have to alternate.

$j(X)$ is prob of choosing that move.

To ~~move~~ transition from C_1 to C_2, generate a ~~random~~ cont. random vector $u^{(1)}$ of length m_1 ⫫ of $\theta^{(1)}$ and let $\theta^{(2)} = f(\theta^{(1)}, u^{(1)})$.

To switch back, do the same with a cont. random vector $u^{(2)}$ of length m_2 ⫫ of $\theta^{(2)}$ and let $\theta^{(1)} = f(\theta^{(2)}, u^{(2)})$

There must be a bijection between $(\theta^{(1)}, u^{(1)})$ and $(\theta^{(2)}, u^{(2)})$, specifically the lengths of $u^{(1)}$ and $u^{(2)}$ must satisfy $n_1 + m_1 = n_2 + m_2$

$q(x, x')$ can now be defined by $u^{(1)}$ and $u^{(2)}$ which have densities q_1 and q_2 with respect to Lebesgue measure in R^{m_1} and R^{m_2}, respectively.

For $A \subset C_1$, $B \subset C_2$ set $\xi(A \times B) = \xi(B \times A) = \lambda\{(\theta^{(1)}, u^{(1)}) : \theta^{(1)} \in A, \theta^{(2)}(\theta, u^{(1)}) \in B\}$

λ is a (n_1+m_1)-dimensional ~~Lebesgue~~ Lebesgue measure.

For general $A, B \subset C$, $\xi(A \times B) = \xi\{(A \cap C_1) \times (B \cap C_2)\} + \xi\{(A \cap C_2) \times (B \cap C_1)\}$.

This ~~achieve~~ achieves symmetry.

For $x = (1, \theta^{(1)}) \in C_1$ and $x' = (2, \theta^{(2)}) \in C_2$,

$$f(x, x') = p(1, \theta^{(1)} | y)\, j(1, \theta^{(1)})\, q_1(u^{(1)}),$$

$$f(x', x) = p(2, \theta^{(2)} | y)\, j(2, \theta^{(2)})\, q_2(u^{(2)}) \left| \frac{\partial(\theta^{(2)}, u^{(2)})}{\partial(\theta^{(1)}, u^{(1)})} \right|$$

and otherwise $f(x, x') = 0$.

Then $\forall x, x' \in C$, $f(x, x')$ is a density with respect to ξ of the equilibrium joint proposal dist. $\pi(dx)P(x, dx')$

$$\underset{x = (1, \theta^{(1)}) \to x' = (2, \theta^{(2)})}{\alpha(x, x')} = \min \left\{ 1, \frac{p(2, \theta^{(2)} | y)\, j(2, \theta^{(2)})\, q_2(u^{(2)})}{p(1, \theta^{(1)} | y)\, j(1, \theta^{(1)})\, q_1(u^{(1)})} \left| \frac{\partial(\theta^{(2)}, u^{(2)})}{\partial(\theta^{(1)}, u^{(1)})} \right| \right\}$$

We can set up the algorithm to where m_1 or $m_2 = 0$

Ex: If $m_2 = 0$, then

$$\alpha(x, x') = \min \left\{ 1, \frac{p(2, \theta^{(2)} | y)\, j(2, \theta^{(2)})}{p(1, \theta^{(1)} | y)\, j(1, \theta^{(1)})\, q_1(u^{(1)})} \left| \frac{\partial(\theta^{(2)})}{\partial(\theta^{(1)}, u^{(1)})} \right| \right\}$$

Modifications can be made.

Coal mining day between 1851 and 1962.

- data $\{y_i, i = 1, 2, \ldots, n\} \in [0, L]$ from a Poisson process with rate given by $x(t)$, log-likelihood is $\sum_{i=1}^{n} \log(x(y_i)) - \int_0^L x(t)\, dt$.

Assume $x(\cdot)$ on $[0, L]$ is a step function

prior:

suppose there are k steps, at $0 < s_1 < s_2 < \ldots < s_k < L$ and that the step function takes a value h_j, the height on a subinterval $[s_j, s_{j+1})$.

for $j = 0, 1, 2, \ldots, k$ (where $s_0 = 0$, $s_{k+1} = L$).

suppose k is drawn from $p(k) = e^{-\lambda} \frac{\lambda^k}{k!}$ conditioned on $k \le k_{max}$.

Given k, $s_1, s_2, \ldots, s_k$ are dist as even-numbered order stats from $2k+1$ points uniformly dist on $[0, L]$

heights $h_0, h_1, \ldots, h_K$ from $\Gamma(\alpha, \beta) = \beta^{\alpha} h^{\alpha-1} e^{-\beta h} / \Gamma(\alpha)$ for $h > 0$.
Can't have $\Gamma(0,0)$. Prior is "close to 'uninformative'".

RJMCMC for step:

4 possible steps:

a) Change the height of a randomly chosen subinterval.
b) change position of a randomly chosen change-point.
c) Create a new change-point in a random location in $[0, L]$. $k \to k+1$
d) remove a randomly chosen change-point. $k \to k-1$

$$\begin{array}{cccc} & & B & D \\ (H, & P, & k, & k-1) \\ \downarrow & \downarrow & \downarrow & \downarrow \\ \eta_k & \pi_k & b_k & d_k \end{array}$$

, which depends on the current k and sum to 1.

$d_0 = \pi_0 = 0$ and $b_{k\,max} = 0$

$b_k = c \min\{1, p(k+1)/p(k)\}$ and $d_{k+1} = c \min\{1, p(k)/p(k+1)\}$

c as large as possible s.t. $b_k + d_k \leq 0.9$
for all $k = 0, 1, \ldots, k_{max}$.

$\rightarrow b_k p(k) = d_{k+1} p(k+1)$

For this ex: $\eta_k = \pi_k$ for $k \neq 0$.

a) If H, then choose $h_0, h_1, \ldots h_K$ at random.
~~To Propose~~ Given h_j propose h_j' s.t. $\log(h_j'/h_j)$ is uniformly dist. on interval $[-\frac{1}{2}, \frac{1}{2}]$

This propsal is by convience.

$$\alpha_H = \min\left\{1, \underbrace{(\text{liklihood ratio})}_{\frac{p(y|x')}{p(y|x)}} \times (h_j'/h_j)^{a} \exp\{-\beta(h_j' - h_j)\}\right\}$$

b) If P, choose $s_1, s_2, \dots, s_k$ at random.
Given s_j, propose s_j' ~~from~~ uniform dist on $[s_{j-1}, s_{j+1}]$

$$\alpha_p = \min\left\{1, (\text{likelihood ratio}) \times \frac{(s_{j+1} - s_j')(s_j' - s_{j-1})}{(s_{j+1} - s_j)(s_j - s_{j-1})}\right\}$$

c) ~~If~~ d) If B, sample uniformly on $[0, L]$ to get s_*
s_* must lie between an existing subinterval (s_j, s_{j+1})

If accepted $s_{j+1}' = s_*$ and $s_{j+1}, s_{j+2}, s_{j+3}, \dots, s_k$

will be relabelled as $s_{j+2}', s_{j+3}', s_{j+4}' \dots s_{k+1}'$

Need new heights for h_j' and h_{j+1}'
for (s_j, s_*) and (s_*, s_{j+1})

h_j is well-supported in post dist. and
so we don't want to completely throw it out.

Thus h_j', h_{j+1}' should be perturbed in either direction from h_j s.t. h_j is a comprimise between them.

To do so $(s_* - s_j)\log(h_j') + (s_{j+1} - s_*)\log(h_{j+1}')$
$= (s_{j+1} - s_j)\log(h_j)$

and define the perturbation $\frac{h_{j+1}'}{h_j} = \frac{1-u}{u}$ where $u \sim U[0,1]$

For D, if we want to conserve detailed balance ~~then removing s_{j+1}~~ the calculation must be reversed. Meaning if s_{j+1} is removed then the new height h_j' from h_j, h_{j+1} must satisfy
$(s_{j+1} - s_j)\log(h_j) + (s_{j+2} - s_{j+1})\log(h_{j+1}) = (s_j' - s_j')\log(h_j')$
where s_{j+1} is drawn at random.

~~$P(y|x) \propto P(x|y)$~~ $P(x'|y)$

For B, $2k+1 \rightarrow 2k+3$, which is accounted for by s_* and u used to separate h'_j, h'_{j+1}.

$$\alpha_B = \min\{1, (\text{likelihood ratio}) \times (\text{prior ratio}) \times (\text{proposal ratio}) \times (\text{Jacobian})\}$$

prior ratio

$$= \min\{1, (\text{likelihood ratio}) \times \frac{P(k+1)}{P(k)} \frac{2(k+1)(2k+3)}{L^2} \frac{(s^* - s_j)(s_{j+1} - s^*)}{s_{j+1} - s_j} * \frac{\beta^\alpha}{\Gamma(\alpha)} \left(\frac{h'_j h'_{j+1}}{h_j}\right)^{\alpha - 1} \times \exp\{-\beta(h'_j + h'_{j+1} - h_j)\}\}$$

$\Downarrow$

recall that the log-likelihood is $\sum_{i=1}^{n} \log\{x(y_i)\} - \int_0^L x(t)\,dt$

$$\text{proposal ratio} = \frac{d_{k+1} L}{b_k (k+1)}$$

$$\text{Jacobian} \quad \frac{(h'_j + h'_{j+1})^2}{h_j}$$

For D, you can just relabel and invert things