**Stat 535 C - Statistical Computing & Monte Carlo Methods**

Lecture 17 - 14th March 2006

Arnaud Doucet

Email: `arnaud@cs.ubc.ca`





- Bayesian Model Selection
- Metropolis-Hastings on a General State-Space
- Trans-dimensional Markov chain Monte Carlo.





- Most Bayesian models discussed until now: prior $p(\theta)$ and likelihood $p(y|\theta)$. Using MCMC, we sample from

$$p(\theta|y) = \frac{p(\theta)\,p(y|\theta)}{\int p(\theta)\,p(y|\theta)\,d\theta}.$$

- We discuss several examples where the model under study is fully specified.

- In practice, we might have a collection of candidate models. This class of problems include cases where "the number of unknowns is something you don't know" (Green, 1995).





- Assume we have a (countable) set $\mathcal{K}$ of candidate models then an associated Bayesian model is such that
  - $k$ denotes the model and has a prior probability $p(k)$
  - $\theta_k \in \Theta_k$ is the unknown parameter associated to model $k$

    of prior $p(\theta_k|k)$.
  - The likelihood is $p(y|k,\theta_k)$.
- You can think of it as a "standard" Bayesian model of parameter $(k,\theta_k) \in \cup_{k\in\mathcal{K}} (\{k\} \times \Theta_k)$.

## 2.1– Bayesian Model Selection

- The Bayes' rule gives the posterior

$$p(k,\theta_k|\,y) = \frac{p(k)\,p(\theta_k|\,k)\,p(y|\,k,\theta_k)}{\sum_{k\in\mathcal{K}}\int_{\Theta_k} p(k)\,p(\theta_k|\,k)\,p(y|\,k,\theta_k)\,d\theta_k}$$

defined on $\cup_{k\in\mathcal{K}}\left(\{k\}\times\Theta_k\right)$.

- From this posterior, we can compute

$$p(k|\,y) \text{ and } \frac{p(y|\,k)}{p(y|\,j)} = \frac{p(k|\,y)}{p(j|\,y)}\frac{p(j)}{p(k)}$$

or performing Bayesian model averaging

$$p(y'|\,y) = \sum_{k\in\mathcal{K}}\int_{\Theta_k} p(y'|\,k,\theta_k)\,p(k,\theta_k|\,y)\,d\theta_k$$



## 2.2– Example: Autoregressive Time Series

- The model $k\in\mathcal{K}=\{1,...,k_{\max}\}$ is given by an AR of order $k$

$$Y_n = \sum_{i=1}^{k} a_i Y_{n-i} + \sigma V_n \text{ where } V_n \sim \mathcal{N}(0,1)$$

and we have $\theta_k = \left(a_{1:k},\sigma^2\right)\in\mathbb{R}^k\times\mathbb{R}^+$.

- We need to defined a prior $p(k,\theta_k) = p(k)\,p(\theta_k|\,k)$, say

$$\begin{aligned} p(k) &= k_{\max}^{-1} \text{ for } k\in\mathcal{K}, \\ p(\theta_k|\,k) &= \mathcal{N}\left(a_{1:k};0,\sigma^2\delta^2 I_k\right)\mathcal{IG}\left(\sigma^2;\frac{\nu_0}{2},\frac{\gamma_0}{2}\right). \end{aligned}$$

- One should be careful, the parameters denoted similarly can have a different "meaning" so that computing say $p\left(\sigma^2\middle|\,y\right)$ does not mean much.

- Some authors favour a more precise notation $\theta_k = \left(a_{k,1:k},\sigma_k^2\right)$ but this can be cumbersome.



## 2.3– Example: Finite Mixture of Gaussians

- The model $k\in\mathcal{K}=\{1,...,k_{\max}\}$ is given by a mixture of $k$ Gaussians

$$Y_n \sim \sum_{i=1}^{k}\pi_i\mathcal{N}\left(\mu_i,\sigma_i^2\right).$$

and we have $\theta_k = \left(\pi_{1:k},\mu_{1:k},\sigma_{1:k}^2\right)\in S_k\times\mathbb{R}^k\times\left(\mathbb{R}^+\right)^k$.

- We need to defined a prior $p(k,\theta_k) = p(k)\,p(\theta_k|\,k)$, say

$$\begin{aligned} p(k) &= k_{\max}^{-1} \text{ for } k\in\mathcal{K}, \\ p(\theta_k|\,k) &= \mathcal{D}(\pi_{1:k};1,...,1)\prod_{i=1}^{k}\mathcal{N}(\mu_i;\alpha,\beta)\,\mathcal{IG}\left(\sigma_i^2;\frac{\nu_0}{2},\frac{\gamma_0}{2}\right). \end{aligned}$$

- Some authors favour a more precise notation $\theta_k = \left(\pi_{k,1:k},\mu_{k,1:k},\sigma_{k,1:k}^2\right)$.



## 2.4– Example: Bayesian Variable Selection

- Assume $Y\in\mathbb{R}$, $X_k\in\mathbb{R}$ and

$$Y = \sum_{\{k:\gamma_k=1\}}\beta_k X_k + \sigma V = \beta_\gamma^{\mathrm{T}} X_\gamma + \sigma V$$

where, for a vector $\gamma = (\gamma_1,...,\gamma_p)$, $\beta_\gamma = \{\beta_k : \gamma_k = 1\}$, $X_\gamma = \{X_k : \gamma_k = 1\}$ and $n_\gamma = \sum_{k=1}^p \gamma_k$.

- Prior distributions

$$\pi_\gamma\left(\beta_\gamma,\sigma^2\right) = \mathcal{N}\left(\beta_\gamma;0,\delta^2\sigma^2 I_{n_\gamma}\right)\mathcal{IG}\left(\sigma^2;\frac{\nu_0}{2},\frac{\gamma_0}{2}\right)$$

and $\pi(\gamma) = \prod_{k=1}^p \pi(\gamma_k) = 2^{-p}$.

- In this case we have $2^p$ models (i.e. configurations of $\gamma$) and the parameter space associated to any vector $\gamma$ is $\mathbb{R}^{n_\gamma}\times\mathbb{R}^+$.

## 3.2– Revisiting the MH algorithm

• We say that a measure $\gamma(dx)$ admits a density with respect to a measure $\lambda(dx)$ if for any (measurable) set $A \in B(\mathcal{X})$

$$\lambda(A) = 0 \Rightarrow \gamma(A) = 0$$

and we call

$$\frac{\gamma(dx)}{\lambda(dx)} = f(x)$$

the density of $\gamma(dx)$ with respect to $\lambda(dx)$.

• In 95% of the applications in statistics $\lambda(dx)$ is the Lebesgue measure $dx$ and we write

$$\frac{\gamma(dx)}{\lambda(dx)} = \frac{\gamma(dx)}{dx} = \gamma(x).$$



## 3.2– Revisiting the MH algorithm

• Consider the STANDARD case where the target is $\pi(dx)$ where $x \in \mathcal{X} \subset \mathbb{R}^d$.

• The MH kernel is given by

$$K(x, dx') = \alpha(x, x')\, q(x, dx') + \left(1 - \int \alpha(x, z)\, q(x, dz)\right) \delta_x(dx')$$

and to ensure its $\pi-$invariance we just to ensure its $\pi-$reversibility

$$\int_{(x,x') \in A \times B} \pi(dx)\, K(x, dx') = \int_{(x,x') \in A \times B} \pi(dx')\, K(x', dx)$$

$$\Leftrightarrow \int_{(x,x') \in A \times B} \pi(dx)\, \alpha(x, x')\, q(x, dx') = \int_{(x,x') \in A \times B} \pi(dx')\, \alpha(x', x)\, q(x', dx)$$

as we always have

$$\int_{(x,x') \in A \times B} \pi(dx) \left(1 - \int \alpha(x, z)\, q(x, dz)\right) \delta_x(dx')$$

$$= \int_{(x,x') \in A \times B} \pi(dx') \left(1 - \int \alpha(x', z)\, q(x', dz)\right) \delta_{x'}(dx)$$



## 3.1– General State-Space Metropolis-Hastings Algorithm

• As stated before, Bayesian model selection problems corresponds to the case where the parameter space is simply $\cup_{k \in \mathcal{K}} (\{k\} \times \Theta_k)$.

• Can we define MCMC algorithms - i.e. Markov chain kernels with fixed invariant distribution $p(k, \theta_k | y)$- ?

• We are going to present a generalization of MH after revisiting first the MH algorithm.



## 3.1– General State-Space Metropolis-Hastings Algorithm

• For such problems, we could use the following approach:
For each $k \in \mathcal{K}$, one could use MCMC to sample from

$$p(\theta_k | y, k) = \frac{p(\theta_k | k)\, p(y | k, \theta_k)}{\int_{\Theta_k} p(\theta_k | k)\, p(y | k, \theta_k)\, d\theta_k} = \frac{p(\theta_k | k)\, p(y | k, \theta_k)}{p(y | k)}.$$

• *Problem*: $\mathcal{K}$ can contain a very large/infinite number of models and many have a very low posterior $p(k | y)$ and so are not relevant for prediction. Moreover, the calculation of $p(y | k)$ is not direct.

## 3.3– Revisiting the Random-Walk Metropolis

• This presentation appears (and is!) unnecessarily complex when $\mathcal{X} \subset \mathbb{R}^d$.

• Assume $x = (x_1, x_2) \in \mathbb{R}^2$ and $u \sim g \in \mathbb{R}$ and we have

$$x_1' = x_1 + u,\ x_2' = x_2,\ u' = -u$$

and we propose the reverse move where $u' \sim g \in \mathbb{R}$

$$x_1 = x_1' + u',\ x_2 = x_2',\ u = -u'$$

and the acceptance probability is simply

$$\alpha\left(x, x'\right) = \min\left(1, \frac{\pi\left(x_1', x_2\right) g\left(x_1 - x_1'\right)}{\pi\left(x_1, x_2\right) g\left(x_1' - x_1\right)}\right)$$

• The main benefit of this approach is that it can also be used whatever the dimension of $x$ in different parts of $\mathcal{X}$ when $\mathcal{X} = \cup_{k \in \mathcal{K}} \left(\{k\} \times \mathbb{R}^{n_k}\right)$.



## 3.2– Revisiting the MH algorithm

• (1)=(2) if (loosely speaking!)

$$\pi\left(x\right) g\left(u\right) \alpha\left(x, x'\right) dx du = \pi\left(x'\right) g'\left(u'\right) \alpha\left(x', x\right) dx' du'$$

• If the transformation $(x, u)$ to $(x', u')$ is a diffeomorphism (the transformation and its inverse are differentiable) then this equality is satisfied if

$$\pi\left(x\right) g\left(u\right) \alpha\left(x, x'\right) = \pi\left(x'\right) g'\left(u'\right) \alpha\left(x', x\right) \left|\frac{\partial\left(x', u'\right)}{\partial\left(x, u\right)}\right|.$$

• It follows that a choice ensuring $\pi$-reversibility is given by

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



## 3.2– Revisiting the MH algorithm

• In practice, we typically define $q\left(x, dx'\right)$ indirectly. Say if $\mathcal{X} \subset \mathbb{R}^d$ then we propose $u \sim g$ of dimension $r$ and then define $x' = h\left(x, u\right)$ so that

$$(1)\ \text{-}\ \int_{(x,x') \in A \times B} \pi\left(dx\right) q\left(x, dx'\right) \alpha\left(x, x'\right) = \int_{(x,x') \in A \times B} \pi\left(x\right) g\left(u\right) \alpha\left(x, x'\right) dx du.$$

• We propose to define the reverse transition by $x = h'\left(x', u'\right)$ where $u' \sim g'$ and

$$(2)\ \text{-}\ \int_{(x,x') \in A \times B} \pi\left(dx'\right) q\left(x', dx\right) \alpha\left(x', x\right) = \int_{(x,x') \in A \times B} \pi\left(x'\right) g'\left(u'\right) \alpha\left(x', x\right) dx' du'.$$

• We want to ensure reversibility i.e. (1)=(2).



## 3.2– Revisiting the MH algorithm

• In the case where we have $\pi\left(dx\right) = \pi\left(x\right) dx$ and $q\left(x, dx'\right) = q\left(x, x'\right) dx'$ and

$$\pi\left(dx\right) \alpha\left(x, x'\right) q\left(x, dx'\right) = \pi\left(dx'\right) \alpha\left(x', x\right) q\left(x', dx\right)$$

$$\Leftrightarrow \pi\left(x\right) \alpha\left(x, x'\right) q\left(x, x'\right) dx dx' = \pi\left(x'\right) \alpha\left(x', x\right) q\left(x', x\right) dx dx'$$

$$\Leftrightarrow \pi\left(x\right) \alpha\left(x, x'\right) q\left(x, x'\right) = \pi\left(x'\right) \alpha\left(x', x\right) q\left(x', x\right)$$

• This is clearly satisfied if

$$\alpha\left(x, x'\right) = \min\left\{1, \frac{\pi\left(x'\right) q\left(x', x\right)}{\pi\left(x\right) q\left(x, x'\right)}\right\} = \min\left\{1, \frac{\pi\left(dx'\right) q\left(x', dx\right)}{\pi\left(dx\right) q\left(x, dx'\right)}\right\}$$

## 4.1– Trans-dimensional MCMC

• Suppose the dimensions of $x, x', u$ and $u'$ are respectively $d, d', r$ and $r'$ then we have functions

$$h : \mathbb{R}^d \times \mathbb{R}^r \to \mathbb{R}^{d'} \quad \text{and} \quad h' : \mathbb{R}^{d'} \times \mathbb{R}^{r'} \to \mathbb{R}^d$$

used respectively for $x' = h(x, u)$ and $x = h'(x', u')$.

• To ensure that we have a diffeomorphism between $(x, u)$ and $(x', u')$, we need the so-called matching condition $d + r = d' + r'$.

• Then we can also used exactly the same reasoning to build the moves.



## 4.2– Example: Birth/Death Moves

• Assume we have a distribution defined on $\{1\} \times \mathbb{R} \cup \{2\} \times \mathbb{R} \times \mathbb{R}$. We want to propose some moves to go from $(1, \theta)$ to $(2, \theta_1, \theta_2)$.

• We can propose $u \sim g \in \mathbb{R}$ and set

$$(\theta_1, \theta_2) = h(\theta, u) = (\theta, u).$$

Its inverse is given by

$$(\theta, u) = h'(\theta_1, \theta_2) = (\theta_1, \theta_2).$$

• The acceptance probability for this "birth" move is given by

$$\min\left(1, \frac{\pi(2, \theta_1, \theta_2)}{\pi(1, \theta)} \frac{1}{g(u)} \left| \frac{\partial(\theta_1, \theta_2)}{\partial(\theta, u)} \right| \right) = \min\left(1, \frac{\pi(2, \theta_1, \theta_2)}{\pi(1, \theta_1) g(\theta_2)}\right).$$



## 4.2– Example: Birth/Death Moves

• The acceptance probability for the associated "death move" is

$$\min\left(1, \frac{\pi(1, \theta)}{\pi(2, \theta_1, \theta_2)} g(u) \left| \frac{\partial(\theta, u)}{\partial(\theta_1, \theta_2)} \right| \right) = \min\left(1, \frac{\pi(1, \theta) g(u)}{\pi(2, \theta, u)}\right)$$

• Once the birth move is defined then the death move follows automatically. In the death move, we do not simulate from $g$ but its expression still appear in the acceptance probability.



## 4.3– Example: Birth/Death Moves

• To simplify notation has in Green (1995) & Robert (2004), we don't emphasize that actually we can have the proposal $g$ which is a function of the current point $\theta$ but it is possible!

• We can propose $u \sim g(\cdot|\theta) \in \mathbb{R}$ and set

$$(\theta_1, \theta_2) = h(\theta, u) = (\theta, u).$$

Its inverse is given by

$$(\theta, u) = h'(\theta_1, \theta_2) = (\theta_1, \theta_2).$$

• The acceptance probability for this "birth" move is given by

$$\min\left(1, \frac{\pi(2, \theta_1, \theta_2)}{\pi(1, \theta)} \frac{1}{g(u|\theta)} \left| \frac{\partial(\theta_1, \theta_2)}{\partial(\theta, u)} \right| \right) = \min\left(1, \frac{\pi(2, \theta_1, \theta_2)}{\pi(1, \theta_1) g(\theta_2|\theta_1)}\right).$$

## 4.3– Example: Birth/Death Moves

• The acceptance probability for the associated "death move" is

$$\min\left(1, \frac{\pi(1,\theta)}{\pi(2,\theta_1,\theta_2)} g(u|\theta) \left|\frac{\partial(\theta,u)}{\partial(\theta_1,\theta_2)}\right|\right) = \min\left(1, \frac{\pi(1,\theta)\, g(u|\theta)}{\pi(2,\theta,u)}\right)$$

• Once the birth move is defined then the death move follows automatically. In the death move, we do not simulate from $g$ but its expression still appears in the acceptance probability.

• Clearly if we have $g(\theta_2|\theta_1) = \pi(\theta_2|2,\theta_1)$ then the expressions simplify

$$\begin{aligned} \min\left(1, \frac{\pi(2,\theta_1,\theta_2)}{\pi(1,\theta_1)\, g(\theta_2|\theta_1)}\right) &= \min\left(1, \frac{\pi(2,\theta_1)}{\pi(1,\theta_1)}\right), \\ \min\left(1, \frac{\pi(1,\theta)\, g(u|\theta)}{\pi(2,\theta,u)}\right) &= \min\left(1, \frac{\pi(1,\theta)}{\pi(2,\theta)}\right). \end{aligned}$$



## 4.4– Example: Split/Merge Moves

• Assume we have a distribution defined on $\{1\} \times \mathbb{R} \cup \{2\} \times \mathbb{R} \times \mathbb{R}$. We want to propose some moves to go from $(1,\theta)$ to $(2,\theta_1,\theta_2)$.

• We can propose $u \sim g \in \mathbb{R}$ and set

$$(\theta_1,\theta_2) = h(\theta,u) = (\theta - u, \theta + u).$$

Its inverse is given by

$$(\theta,u) = h'(\theta_1,\theta_2) = \left(\frac{\theta_1+\theta_2}{2}, \frac{\theta_2-\theta_1}{2}\right).$$

• The acceptance probability for this "split" move is given by

$$\min\left(1, \frac{\pi(2,\theta_1,\theta_2)}{\pi(1,\theta)} \frac{1}{g(u)} \left|\frac{\partial(\theta_1,\theta_2)}{\partial(\theta,u)}\right|\right) = \min\left(1, \frac{\pi(2,\theta_1,\theta_2)}{\pi\left(1,\frac{\theta_1+\theta_2}{2}\right)} \frac{2}{g\left(\frac{\theta_2-\theta_1}{2}\right)}\right).$$



## 4.4– Example: Split/Merge Moves

• The acceptance probability for the associated "merge move" is

$$\min\left(1, \frac{\pi(1,\theta)}{\pi(2,\theta_1,\theta_2)} g(u) \left|\frac{\partial(\theta,u)}{\partial(\theta_1,\theta_2)}\right|\right) = \min\left(1, \frac{\pi(1,\theta)}{\pi(2,\theta-u,\theta+u)} \frac{g(u)}{2}\right)$$

• Once the split move is defined then the merge move follows automatically. In the merge move, we do not simulate from $g$ but its expression still appear in the acceptance probability.



## 4.5– Mixture of Moves

• In practice, the algorithm is based on a combination of moves to move from $x = (k,\theta_k)$ to $x' = (k',\theta_{k'})$ indexed by $i \in \mathcal{M}$ and in this case we just need to have

$$\int_{(x,x')\in A\times B} \pi(dx)\, \alpha_i(x,x')\, q_i(x,dx') = \int_{(x,x')\in A\times B} \pi(dx')\, \alpha_i(x',x)\, q_i(x',dx)$$

to ensure that the kernel $P(x,B)$ defined for $x \notin B$

$$P(x,B) = \sum_{i\in\mathcal{M}} \alpha_i(x,x')\, q_i(x,dx')$$

is $\pi$-reversible.

• In practice, we would like to have

$$P(x,B) = \sum_{i\in\mathcal{M}} j_i(x)\, \alpha_i(x,x')\, q_i(x,dx')$$

where $j_i(x)$ is the probability of selecting the move $i$ once we are in $x$ and $\sum_{i\in\mathcal{M}} j_i(x) = 1$.

## 4.8– Example: Bayesian Model for Autoregressions

• The model $k \in \mathcal{K} = \{1, ..., k_{\max}\}$ is given by an AR of order $k$

$$Y_n = \sum_{i=1}^{k} a_i Y_{n-i} + \sigma V_n \text{ where } V_n \sim \mathcal{N}(0,1)$$

and we have $\theta_k = \left(a_{k,1:k}, \sigma_k^2\right) \in \mathbb{R}^k \times \mathbb{R}^+$ where

$$\begin{aligned} p(k) &= k_{\max}^{-1} \text{ for } k \in \mathcal{K}, \\ p(\theta_k | k) &= \mathcal{N}\left(a_{k,1:k}; 0, \sigma_k^2 \delta^2 I_k\right) \mathcal{IG}\left(\sigma^2; \frac{\nu_0}{2}, \frac{\gamma_0}{2}\right). \end{aligned}$$

• For sake of simplicity, we assume here that the initial conditions $y_{1-k_{\max}:0} = (0, ..., 0)$ are known and we want to sample from

$$p(\theta_k, k | y_{1:T}).$$

• Note that this is not very clever as $p(k | y_{1:T})$ is known up to a normalizing constant!



## 4.7– One minute break

• This brilliant idea is due to P.J. Green, *Reversible Jump MCMC and Bayesian Model Determination*, Biometrika, 1995 although special cases had appeared earlier in physics.

• This is one of the top ten most cited paper in maths and is used nowadays in numerous applications including genetics, econometrics, computer graphics, ecology, etc.



## 4.6– Summary

• For each point $x = (k, \theta_k)$, we define a collection of potential moves selected randomly with probability $j_i(x)$ where $i \in \mathcal{M}$.

• To move from $x = (k, \theta_k)$ to $x' = (k', \theta_{k'})$, we build one (or several) deterministic differentiable and inversible mapping(s)

$$(\theta_{k'}, u_{k',k}) = T_{k,k'}(\theta_k, u_{k,k'})$$

where $u_{k,k'} \sim g_{k,k'}$ and $u_{k',k} \sim g_{k',k}$ and we accept the move with proba

$$\min\left(1, \frac{\pi(k', \theta_{k'}) j_i(k', \theta_{k'}) g_{k',k}(u_{k',k})}{\pi(k, \theta_k) j_i(k, \theta_k) g_{k,k'}(u_{k,k'})} \left| \frac{\partial T_{k,k'}(\theta_k, u_{k,k'})}{\partial(\theta_k, u_{k,k'})} \right| \right).$$



## 4.5– Mixture of Moves

• In this case reversibility is ensured if

$$\int_{(x,x') \in A \times B} \pi(dx) j_i(x) \alpha_i(x,x') q_i(x,dx')$$

$$= \int_{(x,x') \in A \times B} \pi(dx') j_i(x') \alpha_i(x',x) q_i(x',dx)$$

which is satisfied if

$$\alpha_i(x,x') = \min\left(1, \frac{\pi(x') j_i(x') g_i'(u')}{\pi(x) j_i(x) g_i(u)} \left| \frac{\partial(x',u')}{\partial(x,u)} \right| \right).$$

• In practice, we will only have a limited number of moves possible from each point $x$.

## 4.8– Example: Bayesian Model for Autoregressions

- We propose the following moves. If we have $\left(k, a_{1:k}, \sigma_k^2\right)$ then with probability $b_k$ we propose a birth move if $k \leq k_{\max}$, with proba $u_k$ we propose an update move and with proba $d_k = 1 - b_k - u_k$ we propose a death move.

- We have $d_1 = 0$ and $b_{k\max} = 0$.

- The *update move* can simply done in a Gibbs step as

$$p\left(\theta_k \middle| y_{1:T}, k\right) = \mathcal{N}\left(a_{1:k}; m_k, \sigma^2 \Sigma_k\right) \mathcal{IG}\left(\sigma^2; \frac{\nu_k}{2}, \frac{\gamma_k}{2}\right)$$



## 4.8– Example: Bayesian Model for Autoregressions

- *Birth move*: We propose to move from $k$ to $k+1$

$$\left(a_{k+1,1:k}, a_{k+1,k+1}, \sigma_{k+1}^2\right) = \left(a_{k,1:k}, u, \sigma_k^2\right) \text{ where } u \sim g_{k,k+1}$$

and the acceptance probability is

$$\min\left(1, \frac{p\left(a_{k,1:k}, u, \sigma_k^2, k+1 \middle| y_{1:T}\right) d_{k+1}}{p\left(a_{k,1:k}, \sigma_k^2, k \middle| y_{1:T}\right) b_k g_{k,k+1}(u)}\right).$$

- *Death move*: We propose to move from $k$ to $k-1$

$$\left(a_{k-1,1:k-1}, u, \sigma_{k-1}^2\right) = \left(a_{k,1:k-1}, a_{k,k}, \sigma_k^2\right)$$

and the acceptance probability is

$$\min\left(1, \frac{p\left(a_{k,1:k-1}, \sigma_k^2, k-1 \middle| y_{1:T}\right) b_{k-1} g_{k-1,k}\left(a_{k,k}\right)}{p\left(a_{k,1:k}, \sigma_k^2, k \middle| y_{1:T}\right) d_k}\right)$$



## 4.8– Example: Bayesian Model for Autoregressions

- The performance are obviously very dependent on the selection of the proposal distribution. We select whenever possible the full conditional distribution, i.e. we have $u = a_{k+1,k+1} \sim p\left(a_{k+1,k+1} \middle| y_{1:T}, a_{k,1:k}, \sigma_k^2, k+1\right)$ and

$$\begin{aligned} & \min\left(1, \frac{p\left(a_{k,1:k}, u, \sigma_k^2, k+1 \middle| y_{1:T}\right) d_{k+1}}{p\left(a_{k,1:k}, \sigma_k^2, k \middle| y_{1:T}\right) b_k p\left(u \middle| y_{1:T}, a_{k,1:k}, \sigma_k^2, k+1\right)}\right) \\ = & \min\left(1, \frac{p\left(a_{k,1:k}, \sigma_k^2, k+1 \middle| y_{1:T}\right) d_{k+1}}{p\left(a_{k,1:k}, \sigma_k^2, k \middle| y_{1:T}\right) b_k}\right). \end{aligned}$$

- In such cases, it is actually possible to reject a candidate before sampling it!



## 4.8– Example: Bayesian Model for Autoregressions

- We simulate 200 data with $k = 5$ and use 10,000 iterations of RJMCMC.

- The algorithm output is $\left(k^{(i)}, \theta_k^{(i)}\right) \sim p\left(\theta_k, k \middle| y\right)$ (asymptotically).

- The histogram of $\left(k^{(i)}\right)$ yields an estimate of $p\left(k \middle| y\right)$.

- Histograms of $\left(\theta_k^{(i)}\right)$ for which $k^{(i)} = k_0$ yields estimates of $p\left(\theta_{k_0} \middle| y, k_0\right)$.

- The algorithm provides us with an estimate of $p\left(k \middle| y\right)$ which matches analytical expressions.

## 4.9– Summary

- Trans-dimensional MCMC allows us to implement numerically problems with Bayesian model uncertainty.

- Practical implementation is relatively easy, theory behind not so easy...

- Designing efficient trans-dimensional MCMC algorithms is still a research problem.

- On thursday, we will detail several non-trivial examples.