SC7/SM6 Bayes Methods HT20

Lecturer: Geoff Nicholls

Lecture 11: Reversible Jump MCMC I.
(note that these slides were updated shortly after the lecture)

Notes and Problem sheets are available at

http://www.stats.ox.ac.uk/~nicholls/BayesMethods/
What problem does RJMCMC solve?

We begin the lead up to the reversible jump algorithm. What problem does RJMCMC solve? In the joint distribution of the model and the parameter

$$
\pi(\theta, m|y) \propto p(y|\theta, m)\pi(\theta|m)\pi(m), \quad \theta \in \Omega_m, m \in \{1, 2, \ldots, M\}
$$

the dimension of the parameter $\theta$ may vary depending on the model. For eg

$$
m = 1 \quad Y = \alpha + \epsilon \quad \theta = \alpha \quad \Omega_1 = \mathbb{R}
m = 2 \quad Y = \alpha + \beta x + \epsilon \quad \theta = (\alpha, \beta) \quad \Omega_2 = \mathbb{R}^2
$$

and

$$(\theta, m) \in \Omega, \quad \text{where} \quad \Omega = (\Omega_1 \times \{1\}) \cup (\Omega_2 \times \{2\}).$$

“The number of things we dont know is one of things we dont know”. The MCMC algorithm must jump from between spaces of different dimension.
MCMC using proposals on parameterised subspaces

Target probability space \((\Omega, \mathcal{S}, \pi)\) with density \(\pi(d\theta) = \pi(\theta)d\theta\).

Take a density \(g(u), \ u \in \mathcal{U}\) and a function \(\psi_1 : \Omega \times \mathcal{U} \to \Omega\). Given \(\theta\), simulate \(u \sim g(u)\) and set \(\theta' = \psi_1(\theta, u)\). This samples

\[
q(\theta' | \theta) = g(u) \left| \frac{\partial \theta'}{\partial u} \right|^{-1}
\]

(1)

Example: for \(a > 0\) set \(\theta' = \theta + a(2u - 1)\) with \(u \sim \mathcal{U}(0, 1)\),

\[
g(u) = \mathbb{I}_{0 < u < 1}, \quad \psi_1(\theta, u) = \theta + a(2u - 1), \quad q(\theta' | \theta) = \frac{\mathbb{I}_{\theta - a < \theta' < \theta + a}}{2a}.
\]

If \(u' \in \mathcal{U}\) is the \(u\)-value reversing the move then \(u'\) satisfies

\[
\theta = \psi_1(\psi_1(\theta, u), u').
\]

(*)

This determines \(q(\theta' | \theta)\). In the example \(u' = 1 - u\).
We have a mapping between pairs 

\[(\theta', u') = (\psi_1(\theta, u), \psi_2(\theta, u))\]

where \(\psi_2(\theta, u) = u'\) makes (*) work. If \(\psi = (\psi_1, \psi_2)\) then \(\psi\) is a differentiable involution,

\[(\theta, u) = \psi(\psi(\theta, u)).\]

Example: if \(\psi(\theta, u) = (\theta + a(2u - 1), 1 - u)\) as before then

\[\psi(\psi(\theta, u)) = \psi(\theta + a(2u - 1), (1 - u)) = (\theta + a(2u - 1) + a(2(1 - u) - 1), 1 - (1 - u)) = (\theta, u)\]

Up till now we chose \(q(\theta'|\theta)\) and found a density \(g(u)\) and a function \(\theta' = \psi_1(\theta, u)\) to simulate it. Let's just write down \(g\) and \(\psi_1\) and let \(q\) be whatever it is. This gives more freedom.
We are changing variables in a density so we need to consider detailed balance in more detail. Let
\[ K(\theta, d\theta') = \alpha(\theta'|\theta)q(\theta'|\theta)d\theta' + \delta_\theta(d\theta')c(\theta) \] (2)
with \( d\theta' = \left| \frac{\partial \theta'}{\partial u} \right| du \) and
\[ c(\theta) = \left[ 1 - \int_\Omega q(\theta'|\theta)\alpha(\theta'|\theta)d\theta' \right] \]
give the transition kernel \( K \) for the MCMC. The term involving \( c(\theta) \) is the probability for rejection normalising \( K \) over \( \theta' \). DB is
\[ \int_B \int_A \pi(d\theta')K(\theta', d\theta) = \int_A \int_B \pi(d\theta)K(\theta, d\theta'). \] (3)
If this holds for all \( A, B \in S \) then the process is stationary,
\[ \int_B \pi(d\theta') = \int_B \int_\Omega \pi(d\theta)K(\theta, d\theta'). \]
If \( A \cap B = \emptyset \) the term involving \( c(\theta) \) in Eqn 3 is absent, and otherwise cancels on both sides, so DB is equivalent to
\[ \int_B \int_A \pi(\theta')q(\theta|\theta')\alpha(\theta|\theta')d\theta d\theta' = \int_A \int_B \pi(\theta)q(\theta'|\theta)\alpha(\theta'|\theta)d\theta' d\theta. \] (4)
Theorem*: let $\psi : \Omega \times \mathcal{U} \rightarrow \Omega \times \mathcal{U}$ be an invertible, differentiable involution with $\theta, \theta' \in \Omega$ and $u, u' \in \mathcal{U}$.

The MCMC update with proposal $u \sim g(u)$, $(\theta', u') = \psi(\theta, u)$, acceptance ratio

$$r(\theta', u'|\theta, u) = \frac{\pi(\theta')g(u')}{\pi(\theta)g(u)} J_\psi(\theta, u),$$  \hspace{1cm} (5)

with $J_\psi$ the Jacobian for the transformation $(\theta', u') = \psi(\theta, u)$,

$$J_\psi(\theta, u) = \left| \frac{\partial(\theta', u')}{\partial(\theta, u)} \right|,$$

and acceptance probability

$$\alpha(\theta'|\theta) = \min \left\{ 1, r(\theta', u'|\theta, u) \right\}$$  \hspace{1cm} (6)

satisfies DB in Eqn 4 with $q$ given in Eqn 1.

Proof: we need to verify detailed balance, Eqn 4. We begin with
\[ \pi(\theta')q(\theta|\theta')\alpha(\theta|\theta')d\theta d\theta' = \pi(\theta)q(\theta'|\theta)\alpha(\theta'|\theta)d\theta'd\theta. \] (7)
The measures emphasise there is a change of variables.

Change variables in Eqn 7 from \( \theta' \) to \( u \) at fixed \( \theta \). Using Eqn 1,
\[ \pi(\theta)q(\theta'|\theta)\alpha(\theta'|\theta)d\theta'd\theta = \pi(\theta)g(u)\alpha(\theta'(\theta,u)|\theta)dud\theta. \]
Suppose WLOG that \( r(\theta', u'|\theta, u) \leq 1 \). Then the RHS of (7) is
\[
\begin{align*}
\pi(\theta)g(u)\alpha(\theta'(\theta, u)|\theta)dud\theta &= \pi(\theta)g(u)\frac{\pi(\theta')g(u')}{\pi(\theta)g(u)}J_{\psi}(\theta,u)dud\theta \\
&= \pi(\theta')g(u')\left| \frac{\partial(\theta', u')}{\partial(\theta, u)} \right| dud\theta \\
&= \pi(\theta')g(u')du'd\theta'
\end{align*}
\]
since the Jacobian we have is correct for the change of variables.
We assumed $r(\theta', u'|\theta, u) \leq 1$ (WLOG). Suppose it holds that

$$r(\theta', u'|\theta, u) = 1/r(\theta, u|\theta', u'),$$

so that $r(\theta, u|\theta', u') \geq 1$. The LHS of Eqn 4 is then

$$\pi(\theta')q(\theta|\theta')\alpha(\theta|\theta')d\theta d\theta' = \pi(\theta')g(u')du'd\theta'$$

using Eqn 1 with $(\theta', u') \leftrightarrow (\theta, u)$ swapped. DB is satisfied.

We assumed Eqn 8 holds. We now verify this. We have

$$1/r(\theta, u|\theta', u') = \frac{\pi(\theta')g(u')}{\pi(\theta)g(u)}J_{\psi}(\theta', u')^{-1}.$$

and

$$r(\theta', u'|\theta, u) = \frac{\pi(\theta')g(u')}{\pi(\theta)g(u)}J_{\psi}(\theta, u).$$
The standard result for invertible differentiable maps is

\[ J_{\psi^{-1}}(\theta, u) = J_{\psi}(\theta'(\theta, u), u'(\theta, u))^{-1}. \]

when \((\theta', u') = \psi(\theta, u)\). However to get Eqn 8 we need

\[ J_{\psi}(\theta, u) = J_{\psi}(\theta'(\theta, u), u'(\theta, u))^{-1}. \]

This is where \((\theta, u) = \psi(\psi(\theta, u))\) comes in: \(\psi\) is \(\psi^{-1}\).

We verified DB in Eqn 7. We have equality in DB Eqn 4 if the integration domains are equal. This must hold as all the mappings are invertible, so we have the desired result. [EOP]

The Jacobian must be non-singular, so that the change of variables is well defined. The condition \(\text{dim}(\theta', u') = \text{dim}(\theta, u)\) is called “dimension matching”. 
Example: “Random walk on a log scale”. Suppose we are targeting $\theta \sim \exp(1)$ and we use the proposal

$$u \sim U(1/2, 2), \quad \theta' = u\theta \quad \text{so that} \quad (\theta', u') = (u\theta, 1/u)$$

Here $g(u) = 1/(2 - 0.5), 0.5 < u < 2$ and $\dim(\theta', u') = \dim(\theta, u) = 2$ so dimensions match. The Jacobian is $1/u$ since

$$\left| \frac{\partial (\theta', u')}{\partial (\theta, u)} \right| = \left| \begin{array}{cc} u & 0 \\ \theta & -1/u^2 \end{array} \right| = 1/u,$$

The algorithm is as follows. If $X_t = \theta$ then

1) simulate $u \sim U(1/2, 2)$ and set $\theta' = u\theta$;
2) with probability

$$\alpha(\theta'|\theta) = \min \left\{ 1, \frac{\pi(\theta')g(u')}{\pi(\theta)g(u)} \left| \frac{\partial (\theta', u')}{\partial (\theta, u)} \right| \right\} = \min \left\{ 1, u^{-1}e^{\theta'-\theta} \right\},$$

set $X_{t+1} = \theta'$ and otherwise $X_{t+1} = \theta$. 

Factors of $g(u)/g(u')$ cancel in $\alpha$. This proposal is useful if simulating a density which is peaked or diverges at a boundary. This whole framework is a useful generalisation of MCMC.

Exercise: show that the Jacobian for the SRW proposal

$$u \sim U(-a, a), \quad (\theta', u') = (\theta + u, -u)$$

is equal one.

Exercise: in the polynomial regression example in lecture 10, I gave the MCMC updates for $z$ and $\theta$. The update for $\sigma$ is

$$u \sim U(\delta, 1/\delta), \quad \sigma' = u\sigma$$

with $0 < \delta < 1$ a constant we can choose (and adjust for efficient MCMC). Calculate the acceptance probability (answer in L10.R).
Matched proposals

We can allow the proposal density \( g(u) \) generating \( \theta \to \theta' \) to differ from the density \( g'(u') \) generating \( \theta' \to \theta \), so long as

\[
\pi(\theta)K(\theta, \theta')d\theta d\theta' = \pi(\theta')K(\theta', \theta)d\theta' d\theta.
\]

Suppose we carry out MCMC, choosing the \( g(u) \)-update with probability \( \rho \) and the \( g'(u') \)-update wp \( \rho' \).

An acceptance probability achieving DB is

\[
\alpha(\theta'|\theta) = \min \left\{ 1, \frac{\pi(\theta')\rho'g'(u')}{\pi(\theta)\rho g(u)} \left| \frac{\partial(\theta', u')}{\partial(\theta, u)} \right| \right\}.
\]

The proof is essentially the same as before.
Example: target \( \theta \sim \exp(1) \) as before but this time we take \( u \sim U(1, 2) \) wp 1/2 and otherwise \( u \sim U(0.5, 1) \) so \( \rho = \rho' = 1/2 \). The mapping \((\theta', u') = (u\theta, 1/u)\) is the same as before, so the Jacobian does not change. The algorithm becomes

\begin{enumerate}
\item wp 1/2 (a) set \( u \sim U(1, 2) \) else (b) \( u \sim U(0.5, 1) \). Set \( \theta' = u\theta \).
\item if we chose (a) then
\[ g'(u')/g(u) = \frac{U(u'; 0.5, 1)}{U(u; 1, 2)} = 2 \]
and we accept \( \theta' \) wp
\[ \alpha(\theta'|\theta) = \min\{1, 2 \times e^{-\theta'+\theta}-\log(u)\} \],
and if we chose (b) then \( g'(u')/g(u) = 0.5 \) and we accept \( \theta' \) wp
\[ \alpha(\theta'|\theta) = \min\{1, 0.5 \times e^{-\theta'+\theta}-\log(u)\} \].
\end{enumerate}