Avd. Matematisk statistik

KTH Teknikvetenskap

Sf2955 COMPUTER INTENSIVE METHODS
MCMC
On the Reversible Jump Monte Carlo Markov Chain
2009
Timo Koski

1 Introduction

1.1 Model choice

MCMC methods for (Bayesian) computation have so far in these lectures been restricted to to cases, where the dimensionality of the domain of the target distribution has been fixed. There are a number of problems involving inference about curves, surfaces and/or images, where the dimension of the object is not fixed. Some examples involve model choice; including

1. variable selection in regression
2. estimation of the number of components in a mixture
3. estimating the order of ARMA(p, q)
4. detecting the number of change points (Gustafsson, 2000, p.246)

In this lecture we study the construction of Markov chain samplers that jump between the parameter subspaces of differing dimensionality. The technique is called *Reversible jump Markov chain Monte Carlo* and is due to (Green 1995, see also Grenander and Miller 1994). This lecture follows mainly the tutorial (Waagepetersen and Sorenson, 2001) and (Sorensen and Gianola 2002) adding a few details from (Green 1995).

1.2 On the Formalism

The notations in the sequel are slightly involved, since we must take into account for the fact that parameters in general may change dimension as the Markov chain jumps from one model to another.

2 Bayesian model choice via a hierarchical model

2.1 Notation

2.1.1 Models, Nested Models

We look at a countable number of candidate models $\{\mathcal{M}_k, k \in \mathcal{K}\}$. The model $\mathcal{M}_k$ has a vector of unknown parameters, denoted by $\theta^{(k)}$, which is an n_k-dimensional vector. The dimension of $\theta^{(k)}$ may vary from model to model.

For given k

$$\left(k, \theta^{(k)}\right) \in \mathcal{C}_k := \{k\} \times \Theta^{n_k},$$

where $\dim(\Theta) = n_k$, or

$$\left(k, \theta^{(k)}\right) \in \mathcal{C}_k := \{k\} \times \mathcal{R}^{n_k}.$$

Example 2.1 We consider an example with two models $\mathcal{K} = \{1, 2\}$. $\mathcal{C}_1 = \{1\} \times \mathcal{R}$, $\theta^{(1)} = \theta$, $\mathcal{C}_2 = \{2\} \times \mathcal{R}^2$, $\theta^{(2)} = (\theta_1, \theta_2)$.

These might be the constant in noise

$$\mathcal{M}_1 : \quad y = \theta + \epsilon,$$

where

$$\theta^{(1)} = \theta$$

and the linear regression

$$\mathcal{M}_2 : \quad y = \theta_1 + \theta_2 x + \epsilon, ,$$

so that

$$\theta^{(2)} = (\theta_1, \theta_2) .$$

The example is to be continued ...

■

Two models $\mathcal{C}_k$ and $\mathcal{C}_{k'}$ are called **nested**, if

$$\mathcal{C}_k \subset \mathcal{C}_{k'}.$$

If we take $\theta \leftrightarrow \theta_1, \theta_2 = 0$ in example 2.1, we can regard $\mathcal{M}_1$ as being nested in $\mathcal{M}_2$, linear regression. Another similar case is the estimation of sinusoids in white Gaussian noise (Andrieu and Doucet, 1999).

2.1.2 Notation for Hierarchic Bayesian Inference

We observe data y. The hierarchical structure is expressed by modelling the joint distribution of $\left(k, \theta^{(k)}, y\right)$ as

$$p\left(k, \theta^{(k)}, y\right) = l\left(y|\theta^{(k)}, k\right) \psi\left(\theta^{(k)}|k\right) pr\left(k\right), \tag{2.1}$$

which is the product of likelihood (l), prior on parameters (ψ), and prior (pr) on the models $\mathcal{C}_k$, respectively.

Bayesian inference is based on the joint posterior

$$p\left(k, \theta^{(k)} \mid y\right) = p\left(\theta^{(k)} \mid y, k\right) p\left(k \mid y\right).$$

We shall now construct a Markov chain $\{X_n\}_{n \geq 0}$ with $\{p\left(k, \theta^{(k)} \mid y\right)\}_{k \in \mathcal{K}}$ as the target distribution. If k was fixed, we could apply the usual Metropolis-Hastings algorithm.

The problem in implementing an MCMC algorithm is the need to be able to move from a model $\mathcal{C}_k$ to another model $\mathcal{C}_{k'}$, where $k < k^{'}$ poses a particular difficulty.

Let us set

$$x = \left(k, \theta^{(k)}\right).$$

The chain is implemented by random choices between available **moves** at each transition in order to traverse freely across the combined parameter space of the model family

$$\mathcal{C} = \bigcup_{k \in \mathcal{K}} \mathcal{C}_k,$$

i.e., $x \in \mathcal{C}$. We need a Markov chain that attains the detailed balance condition within each move type.

3 The reversible jump sampler

3.1 Notations

Let us introduce

$$X_n := (M_n, Z_n),$$

where M_n assumes values in $\mathcal{K}$. Given $M_n = k$, Z_n takes values in $\mathcal{R}^{n_k}$. We set

$$X_n = x \Leftrightarrow \left(M_n = k, Z_n = \theta^{(k)}\right).$$

Then, if $A \subseteq \mathcal{R}^{n_k}$, $\theta^{(k)} = z$

$$P(M_n = k, Z_n \in A) = p(k|y) \int_A p(z \mid y, k)\, dz$$

$$= p(k|y) \int_A p_k(z \mid y)\, dz.$$

Hence $p_k(z \mid y)$ is posterior density of $\theta^{(k)}$ given the model $\mathcal{M}_k$ and the data y, and $p(k|y)$ is the posterior probability of the model $\mathcal{M}_k$ given the data y. Hence

$$\{p(k|y)\, p_k(z|y)\}_{(k,z)\in\mathcal{C}}$$

is the target distribution in this context. Next, we get from (2.1)

$$\begin{aligned} p(k|y)\, p_k(z|y) &= p(k, z \mid y) = \frac{p(k,z,y)}{p(y)} \\ &= C^{-1} l(y|z,k)\, \psi(z|k)\, pr(k), \end{aligned} \tag{3.1}$$

where

$$C = \sum_{k\in\mathcal{K}} pr(k) \int \psi(z \mid k)\, dz. \tag{3.2}$$

To see the conceptual difficulties in finding the *detailed balance* (c.f. previous lecture) in this setting, let us write

$$p(x \mid y) = p\left(k, \theta^{(k)} \mid y\right), x \in \mathcal{C}.$$

Hence, for $x \neq x'$ the densities $p(x \mid y)$ and $p\left(x' \mid y\right)$ may be defined in spaces of different dimension. Suppose, as in the example 2.1 above,

$$x = \left(k, \theta^{(1)}\right), x' = \left(k', \theta^{(2)}\right), \quad n_1 = 1 < n_2 = 2.$$

Then $p(\theta \mid y, k)$ is a density on $\mathcal{R}$ and $p\left(\theta^{(1)}, \theta^{(2)} \mid y, k^{'}\right)$ is a density on $\mathcal{R}^2$. Technically formulated , when $\mathcal{R}$ is imbedded as a subspace in $\mathcal{R}^2$, the density $p(\theta \mid y, k)$ corresponds thus to a measure, which is singular w.r.t. the measure corresponding to $p\left(\theta^{(1)}, \theta^{(2)} \mid y, k^{'}\right)$.

3.2 Dimension Matching Transformation

The proposals are generated by construction of a dimension matching transformation, as explained next. Suppose that the current state is

$$X_n = (k, z)$$

and that we consider a move to

$$\left(k^{'}, z^{'}\right)$$

also expressed as

$$k \mapsto k^{'}, z \mapsto z^{'}.$$

A proposal is written as

$$Y_{n+1} = \left(Y_{n+1}^{ind}, Y_{n+1}^{par}\right), \tag{3.3}$$

where the superscript *ind* is a label for the proposed model M_{n+1} and *par* is a label for the proposal Z_{n+1}.

The proposal is accepted with probability

$$\alpha_{k \mapsto k'}\left(z, Y_{n+1}^{par}\right).$$

The probability $\alpha_{k \mapsto k'}$ is derived in section (4.5) below.

It is easy to generate $M_{n+1} = k^{'}$ by a *single component updating*. We define a probability transition matrix $\left\{p_{i|j}\right\}_{i,j \in \mathcal{K} \times \mathcal{K}}$ and draw a value $k^{'}$ from the row $\{p_{k|j}\}_{j \in \mathcal{K}}$.

Given $Y_{n+1}^{ind} = k^{'}$, we take a function $g_{1k \mapsto k'}$ such that

$$Y_{n+1}^{par} = g_{1k \mapsto k'}(z, U),$$

where U is a random vector, The dimension of the vector U is denoted by (the strange symbol) $n_{k \mapsto k'}$. U has a (proposal) density

$$q_{k \mapsto k'}(z, \cdot) \tag{3.4}$$

on $\mathcal{R}^{n_{k \mapsto k'}}$. The dimension $n_{k \mapsto k'}$ is to be adjusted to the dimension required by the reverse move, which is known as *dimension matching*.

3.3 Dimension matching

Consider the move from (k, z) to

$$\left(k^{'}, z^{'}\right) = \left(k^{'}, g_{1k\mapsto k'}\left(z, U\right)\right),$$

and a move in the opposite direction from $\left(k^{'}, z^{'}\right)$ to

$$(k, z) = \left(k^{'}, g_{1k'\mapsto k}\left(z^{'}, U^{'}\right)\right),$$

where $U^{'}$ is a random vector, which has a (proposal) density

$$q_{k'\mapsto k}(z^{'}, \cdot) \tag{3.5}$$

Further it will be assumed that the functions $g_{2k\mapsto k'}$ and $g_{2k'\mapsto k}$ are such that the mapping $g_{k\mapsto k'}$ given by

$$\left(z^{'}, U^{'}\right) = g_{k\mapsto k'}(z, U) = \left(g_{1k\mapsto k'}(z, U), g_{2k\mapsto k'}(z, U)\right), \tag{3.6}$$

is one-to-one (invertible) with

$$\begin{aligned}(z, U) &= g^{-1}_{k\mapsto k'}(z^{'}, U^{'}) \\ = g_{k'\mapsto k}\left(z^{'}, U^{'}\right) &= \left(g_{1k'\mapsto k}(z^{'}, U^{'}), g_{2k'\mapsto k}(z^{'}, U^{'})\right),\end{aligned} \tag{3.7}$$

and that $g_{k\mapsto k'}$ is differentiable.

The transformations (3.6) and (3.7) are possible, if we can construct the mappings $g_{k\mapsto k'}$ and $g_{k'\mapsto k}$ to be one-to-one; a necessary condition for the existence of a one-to-one mapping is that the vectors

$$(z, U)$$

and

$$\left(z^{'}, U^{'}\right)$$

must have the same dimension. This imposes the **dimension matching condition**

$$n_k + n_{k\mapsto k'} = n_{k'} + n_{k'\mapsto k}. \tag{3.8}$$

With regard to dimension matching we can distinquish **three types of moves** depending on the dimension of U. Let us assume $n_k < n_{k'}$.

a) $n_{k\mapsto k'} = n_{k'} - n_k$

b) $n_{k'} - n_k < n_{k\mapsto k'} < n_{k'}$

c) $n_{k'} < n_{k\mapsto k'}$

The class a) moves are most commonly used for transitions between nested models, where we add or delete parameters from the current model to the next.

If $n_{k\mapsto k'}$, or $n_{k'\mapsto k}$, or both are equal to zero, then we talk about **deterministic moves between models**.

Example 3.1 *[Example 2.1 continued]* We choose to move from $\mathcal{C}_1 = \{1\} \times \mathcal{R}$ to $\mathcal{C}_2 = \{2\} \times \mathcal{R}^2$ by the map

$$(\theta_1, \theta_2) = g_{1\mapsto 2}(\theta, U) = (\theta - U, \theta + U),$$

where U is a real random variable. Equivalently

$$g_{1\mapsto 2}(\theta, U) = (g_{11\mapsto 2}(\theta, U), g_{21\mapsto 2}(\theta, U)).$$

We move from $\mathcal{C}_2 = \{2\} \times \mathcal{R}^2$ to $\mathcal{C}_1 = \{1\} \times \mathcal{R}$ by

$$\theta = \frac{1}{2}(\theta_1 + \theta_2),$$

which is clearly a deterministic move. We can express this as

$$(\theta, U) = g_{2\mapsto 1}(\theta_1, \theta_2) = \left(\frac{1}{2}(\theta_1 + \theta_2), \frac{1}{2}(\theta_2 - \theta_1)\right)$$
$$\Leftrightarrow$$
$$g_{2\mapsto 1}\left(\theta', U'\right) = \left(g_{12\mapsto 1}\left(\theta', U'\right), g_{22\mapsto 1}\left(\theta', U'\right)\right).$$

The dimension matching criterion is thus

$$\begin{aligned} n_k + n_{k\mapsto k'} &= n_1 + n_{1\mapsto 2} \\ &= 1 + 1 && (3.9) \\ &= 2 + 0 = n_2 + n_{2\mapsto 1} = n_{k'} + n_{k'\mapsto k}. && (3.10) \end{aligned}$$

We can write the preceding as

$$\begin{pmatrix} \theta_1 \\ \theta_2 \end{pmatrix} = \begin{pmatrix} 1 & -1 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} \theta \\ U \end{pmatrix} = A \begin{pmatrix} \theta \\ U \end{pmatrix} \qquad (3.11)$$

and

$$\begin{pmatrix} \theta \\ U \end{pmatrix} = \begin{pmatrix} \frac{1}{2} & \frac{1}{2} \\ -\frac{1}{2} & \frac{1}{2} \end{pmatrix} \begin{pmatrix} \theta_1 \\ \theta_2 \end{pmatrix} = A^{-1} \begin{pmatrix} \theta_1 \\ \theta_2 \end{pmatrix}$$

In this example $g_{k \mapsto k'}$ ($= g_{1 \mapsto 2}$) and $g_{k' \mapsto k}$ ($= g_{2 \mapsto 1}$) are linear, invertible, and differentiable. The example is to be continued.

■

3.4 The proposal kernel

We define in general the proposal kernel using (3.4) as

$$\begin{aligned} Q_{k \mapsto k'}(z, B_{k'}) &:= P\left(Y_{n+1}^{ind} = k', Y_{n+1}^{par} \in B_{k'} | X_n = (k, z)\right) \\ &= p_{k|k'} \int_{\{u | g_{1k \mapsto k'}(z,u) \in B_{k'}\}} q_{k \mapsto k'}(z, u) du, \end{aligned} \tag{3.12}$$

and with (3.5)

$$\begin{aligned} Q_{k' \mapsto k}\left(z', A_k\right) &:= P\left(Y_{n+1}^{ind} = k, Y_{n+1}^{par} \in A_k | X_n = (k', z')\right) \\ &= p_{k'|k} \int_{\{u' | g_{1k' \mapsto k}\left(z', u'\right) \in A_k\}} q_{k' \mapsto k}(z', u') du'. \end{aligned} \tag{3.13}$$

Example 3.2 *[Example 3.1 continued, more on dimension matching]* In this example we have in (3.12) with $B_2 \subset \mathcal{R}^2$

$$Q_{1 \mapsto 2}(\theta, B_2) = p_{1|2} \int_{\{u | (\theta - u, \theta + u) \in B_2\}} q_{1 \mapsto 2}(\theta, u) du.$$

The reverse move is deterministic, so that in (3.13) for $A_1 \subset \mathcal{R}$,

$$Q_{2 \mapsto 1}((\theta_1, \theta_2), A_1) = \begin{cases} p_{2|1} & \text{if } \frac{\theta_1 + \theta_2}{2} \in A_1 \\ 0 & \text{otherwise.} \end{cases}$$

■

4 Reversibility and the acceptance probability

4.1 An Outline

The derivation of the acceptance probability of moves between spaces is presented by starting from the reversibility condition.

- We express the transition kernel in terms of a proposal kernel and an acceptance probability, and find the condition for detailed balance.
- Second, a change of variable is performed that allows both sides of the detailed balance to be expressed in terms of the same parameters. This requires monotonity of the transformations ($g_{k \mapsto k'}$ and $g_{k' \mapsto k}$) between the parameters, or that the dimension matching condition holds. Identification of the conditions for equality between the probability of opposite moves, and for detailed balance to hold, leads to the final step.

4.2 Reversibility Condition

We need to investigate the consequences of the condition

$$\begin{aligned} & P\left(M_n = k, Z_n \in A_k, M_{n+1} = k^{'}, Z_{n+1} \in B_{k'}\right) \\ & = P\left(M_n = k^{'}, Z_n \in B_{k'}, M_{n+1} = k, Z_{n+1} \in A_k\right) \end{aligned} \tag{4.1}$$

for all $k, = k^{'}$ and $A_k, B_{k'}$. We write the left hand side of (4.1) as

$$P\left(M_n = k, Z_n \in A_k, M_{n+1} = k^{'}, Z_{n+1} \in B_{k'}\right)$$

$$= p\left(k|y\right) \int_{A_k} P\left(M_{n+1} = k^{'}, Z_{n+1} \in B_{k'} | M_n = k, Z_n = z\right) p_k\left(z|y\right) dz, \tag{4.2}$$

since $p\left(k|y\right) p_k\left(z|y\right)$ is the target density on $\mathcal{C}$, and therefore plays the role of an invariant density. Here

$$P\left(M_{n+1} = k^{'}, Z_{n+1} \in B_{k'} | M_n = k, Z_n = z\right)$$

is the transition kernel, to be derived next.

We define first

$$Q^{\mathbf{a}}_{k\mapsto k'}\left(z,B_{k'}\right):=P\left(Y^{ind}_{n+1}=k^{'},Y^{par}_{n+1}\in B_{k'},\text{and }Y_{n+1}\text{ is accepted }|X_n=(k,z)\right).$$

We introduce also

$$s_k(z)=P\left(Y_{n+1}\text{ is rejected }|X_n=(k,z)\right),$$

as the conditional probability of rejecting the proposal given that $X_n=(k,z)$. Then the transition kernel is written, like in the preceding lecture, as

$$P\left(M_{n+1}=k^{'},Z_{n+1}\in B_{k'}|M_n=k,Z_n=z\right)=$$
$$=Q^{\mathbf{a}}_{k\mapsto k'}\left(z,B_{k'}\right)+I_{k',B_{k'}}(k,z)s_k(z).$$

Note that it can happen that $Z_{n+1}\in B_{k'}$ even if the proposal was rejected, in case $k=k^{'}$ and $z\in B_{k'}$.

When we insert this in (4.2) we get

$$p\left(k|y\right)\int_{A_k}P\left(M_{n+1}=k^{'},Z_{n+1}\in B_{k'}|M_n=k,Z_n=z\right)p_k\left(z|y\right)dz$$

$$\begin{aligned}&=p\left(k|y\right)\int_{A_k}Q^{\mathbf{a}}_{k\mapsto k'}\left(z,B_{k'}\right)p_k\left(z|y\right)dz\\&+p\left(k|y\right)\int_{A_k}I_{k',B_{k'}}(k,z)s_k(z)p_k\left(z|y\right)dz.\end{aligned}\tag{4.3}$$

By symmetry, the right hand side of (4.1) equals

$$\begin{aligned}&=p\left(k^{'}|y\right)\int_{B_{k'}}Q^{\mathbf{a}}_{k'\mapsto k}\left(z^{'},A_k\right)p_k\left(z^{'}|y\right)dz^{'}\\&+p\left(k|y\right)\int_{B_{k'}}I_{k',A_k}(k^{'},z^{'})s_{k'}(z^{'})p_k\left(z^{'}|y\right)dz^{'}.\end{aligned}\tag{4.4}$$

Here, we compare the last terms in (4.3) and (4.4)

$$p\left(k|y\right)\int_{A_k}I_{k',B_{k'}}(k,z)s_k(z)p_k\left(z|y\right)dz$$

$$= p\,(k|y) \int_{B_{k'}} I_{k',A_k}(k^{'}, z^{'}) s_{k'}(z^{'}) p_k \left(z^{'}|y\right) dz^{'},$$

since if $k \neq k^{'}$, the indicator fuctions are both equal to zero, and if $k = k^{'}$, the move is inside the same model, and the expressions are identical.

Therefore, from (4.3) and (4.4), if

$$p\,(k|y) \int_{A_k} Q^{\mathbf{a}}_{k \mapsto k'}\,(z, B_{k'})\, p_k\,(z|y)\, dz$$
$$= p\left(k^{'}|y\right) \int_{B_{k'}} Q^{\mathbf{a}}_{k' \mapsto k}\left(z^{'}, A_k\right) p_{k'}\left(z^{'}|y\right) dz^{'} \tag{4.5}$$

then the desired reversibility property holds.

4.3 A More Detailed Expression

We shall write the equation (4.5) in a more explicit form. Let us recall that

a) Y^{ind}_{m+1} is drawn from $p_{k|k'}$

b) Y^{par}_{m+1} is generated in $\mathcal{C}_{k'}$ and belongs to $B_{k'}$

$$\Leftrightarrow g_{1k \mapsto k'}(z, U) = z^{'} \in B_{k'}.$$

c) Y_{n+1} is accepted with probability $\alpha_{k \mapsto k'}\,(z, g_{1k \mapsto k'}(z, U))$, or

$$\alpha_{k \mapsto k'}\,(z, g_{1k \mapsto k'}(z, U)) = \alpha_{k \mapsto k'} \left(z, z^{'}\right),$$

and

$$U \sim q_{k \mapsto k'}(z, \cdot). \tag{4.6}$$

Taking a) - c) above into account, we get

$$Q^{\mathbf{a}}_{k \mapsto k'}\,(z, B_{k'}) =$$
$$= p_{k|k'} \int I_{B_{k'}}\left(z^{'}\right) \alpha_{k \mapsto k'}\left(z, z^{'}\right) q_{k \mapsto k'}(z, u) du.$$

If we insert this in the left hand side of (4.5), we have

$$p\,(k|y) \int_{A_k} Q^{\mathbf{a}}_{k \mapsto k'}\,(z, B_{k'})\, p_k\,(z|y)\, dz$$

$$= p\left(k|y\right) \int_{A_k} p_{k|k'} \int I_{B_{k'}} \left(z^{'}\right) \alpha_{k\mapsto k'} \left(z, z^{'}\right) q_{k\mapsto k'}(z,u) du p_k\left(z|y\right) dz$$

$$= p\left(k|y\right) \int p_{k|k'} \int I_{A_k, B_{k'}} \left(z, z^{'}\right) \alpha_{k\mapsto k'} \left(z, z^{'}\right) q_{k\mapsto k'}(z,u) p_k\left(z|y\right) du dz.$$

By symmetry, we have in the right hand side of (4.5),

$$p\left(k^{'}|y\right) \int_{B_{k'}} Q^{\mathbf{a}}_{k'\mapsto k} \left(z^{'}, A_k\right) p_{k'} \left(z^{'}|y\right) dz^{'}$$

$$= p\left(k^{'}|y\right) \int\int I_{B_{k'}, A_k} \left(z^{'}, z\right) p_{k'|k} \alpha_{k'\mapsto k} \left(z^{'}, z\right) q_{k'\mapsto k}(z^{'}, u^{'}) p_{k'} \left(z^{'}|y\right) dz^{'} du^{'}.$$

4.4 Change of Variable

Hence we require the following equality to hold

$$\begin{aligned} p\left(k|y\right) \int p_{k|k'} \int I_{A_k, B_{k'}} \left(z, z^{'}\right) \alpha_{k\mapsto k'} \left(z, z^{'}\right) q_{k\mapsto k'}(z,u) p_k\left(z|y\right) dz du \\ = \\ p\left(k^{'}|y\right) \int\int I_{B_{k'}, A_k} \left(z^{'}, z\right) p_{k'|k} \alpha_{k'\mapsto k} \left(z^{'}, z\right) q_{k'\mapsto k}(z^{'}, u^{'}) p_{k'} \left(z^{'}|y\right) dz^{'} du^{'}. \end{aligned} \tag{4.7}$$

Because the maps $g_{k\mapsto k'}$ and $g_{k'\mapsto k}$ in (3.6) and (3.7) are invertible, we may make the change of variable

$$\left(z^{'}, u^{'}\right) = g_{k\mapsto k'}(z,u).$$

The Jacobian of the transformation is

$$|\det \frac{\partial g_{k\mapsto k'}(z,u)}{\partial(z,u)}|.$$

Then the standard rule of transformation of variables in a multiple integral yields in the right hand side of (4.7)

$$\begin{aligned} p\left(k^{'}|y\right) \int\int I_{B_{k'}, A_k} \left(z^{'}, z\right) p_{k'|k} \alpha_{k'\mapsto k} \left(z^{'}, z\right) q_{k'\mapsto k}(z^{'}, u^{'}) p_{k'} \left(z^{'}|y\right) dz^{'} du^{'} \\ = \\ p\left(k^{'}|y\right) \int\int I_{B_{k'}, A_k} \left(z^{'}, z\right) p_{k'|k} \alpha_{k'\mapsto k} \left(z^{'}, z\right) q_{k'\mapsto k}(z^{'}, u) p_{k'} \left(z^{'}|y\right) |\det \frac{\partial g_{k\mapsto k'}(z,u)}{\partial(z,u)}| dz du. \end{aligned} \tag{4.8}$$

Thus there is an inequality between the left hand side of (4.7) and the expression in the right hand side of (4.8), if

$$p\left(k|y\right)p_{k|k'}\alpha_{k\mapsto k'}\left(z,z'\right)q_{k\mapsto k'}(z,u)p_k\left(z|y\right)$$
$$= \tag{4.9}$$
$$p\left(k'|y\right)p_{k'|k}\alpha_{k'\mapsto k}\left(z',z\right)q_{k'\mapsto k}\left(z',u'\right)p_{k'}\left(z'|y\right)|\det\frac{\partial g_{k\mapsto k'}(z,u)}{\partial(z,u)}|.$$

4.5 Acceptance Probability

Clearly, the last equality in (4.9) above holds if

$$\alpha_{k\mapsto k'}\left(z,z'\right)= \tag{4.10}$$
$$\min\left(1,\frac{p\left(k'|y\right)p_{k'|k}q_{k'\mapsto k}\left(z',u'\right)p_{k'}\left(z'|y\right)}{p\left(k|y\right)p_{k|k'}q_{k\mapsto k'}(z,u)p_k\left(z|y\right)}|\det\frac{\partial g_{k\mapsto k'}(z,u)}{\partial(z,u)}|\right).$$

Example 4.1 *[Example 3.1 continued]* In the example 3.1 we have

$$|\det\frac{\partial g_{1\mapsto 2}(\theta,u)}{\partial(\theta,u)}|=\det A=2,$$

where the matrix A is given in (3.11), and

$$|\det\frac{\partial g_{2\mapsto 1}\left(\theta_1,\theta_2\right)}{\partial\left(\theta_1,\theta_2\right)}|=\det A^{-1}=\frac{1}{2}.$$

■

After having glanced at (4.10), let us make three remarks:

1. Note that if the move is inside a model, i.e., $k=k'$, the acceptance probability above reduces to

$$\min\left(1,\frac{q_{k'\mapsto k}\left(z',u'\right)p_{k'}\left(z'|y\right)}{q_{k\mapsto k'}(z,u)p_k\left(z|y\right)}\right),$$

   which is the acceptance probability in a Metropolis-Hastings chain (Hastings 1970).

2. In view of (3.1) we have that

$$p(k|y)\,p_k(z|y) = C^{-1} l(y|z,k)\,\psi(z|k)\,pr(k)\,.$$

Thus we do not need to know the constant C in (3.2) to implement the reversible jump sampler. However, the normalizing constant in ψ may depend on k, and will then have to be known.

3. The crucial problem in reversible jump MCMC is to find a transformation $g_{k \mapsto k'}$, which is invertible and differentiable, as well as to find the the proposal density $q_{k \mapsto k'}$. Rules for these efforts are developed in (Brooks et.al. 2003).

   An additional problem is to compute the Jacobian[1].

5 Stochastic Stability

Finally, we should prove that the reversible jump Markov chain as constructed above is aperiodic, and ϕ-irreducible, and that it converges to the invariant distribution. Here the general theory of Markov chains is needed. One such proof is presented in (Andrieu and Doucet, 1999) for the choice of sinusoids in white Gaussian noise.

6 An Exercise

This exercise is concerned with model choice between a gamma distributed distribution and a lognormal distribution. The gamma distribution is defined by the probability density

$$l(y|\alpha,\beta,k=1) = \frac{1}{\Gamma(\alpha)} y^{\alpha-1} e^{-y/\beta}, \quad 0 < y < \infty, \alpha > 0, \beta > 0. \tag{6.11}$$

The lognormal distribution is defined by the the probability density

$$l\left(y|\mu,\sigma^2,k=2\right) = \frac{1}{y\sqrt{2\pi\sigma}} e^{-(\ln y-\mu)^2/\sigma^2}, \quad 0 < y < \infty, -\infty < \mu < \infty, \sigma > 0. \tag{6.12}$$

[1] To quote Robert and Casella (1999, p. 261): *..nothing is so easy as to write down the wrong Jacobian..*

In this case both models have the same number of parameters, and thus dimension matching allows us to use deterministic moves, if we so desire.

Then the posterior probability for the gamma model is

$$p(1, \alpha, \beta \mid y) \propto l(y|\alpha, \beta, k = 1)\, \psi(\alpha, \beta|1)\, pr(1)\,.$$

where $\psi(\alpha, \beta|1)$ is a prior density for the parameters of the gamma density. For the lognormal model we have

$$p(2, \mu, \sigma \mid y) \propto l\left(y|\mu, \sigma^2, k = 2\right) \psi\left(\mu, \sigma^2|2\right) pr(2)\,.$$

where $\psi(\mu, \sigma^2|2)$ is a prior density for the parameters of the lognormal density.

To accomodate with our notations above we set

$$(k, z) = (1, (\alpha, \beta))$$

and

$$(k', z') = \left(2, \left(\mu, \sigma^2\right)\right),$$

and we consider the move from (k, z) to (k', z'). One way to propose values for (μ, σ^2) would be to equate the first and second order moments under gamma and lognormal models to each other. This is a deterministic move, and yields

$$\begin{aligned} \mu &= \ln\left(\frac{\alpha\beta}{\sqrt{1+\frac{1}{\alpha}}}\right) \\ \sigma^2 &= \ln\left(1 + \frac{1}{\alpha}\right). \end{aligned} \tag{6.13}$$

These equations define in other words

$$\left(\mu, \sigma^2\right) = g_{1\mapsto 2}(\alpha, \beta)\,.$$

The **homework assignments** are:

a) Derive the equations in (6.13).

b) Find

$$(\alpha, \beta) = g_{2\mapsto 1}\left(\mu, \sigma^2\right).$$

3) Show that

$$|\det \frac{\partial g_{1\mapsto 2}(\alpha,\beta)}{\partial(\alpha,\beta)}| = \frac{1}{\alpha\beta(\alpha+1)},$$

4) Implement a reversible jump MCMC for choice between gamma and lognormal models using pseudo-random numbers y drawn from a gamma distribution. One way to do the simulation is to use `gamrnd` in Matlab^{R} Statistics Toolbox. Show the posterior density (in some way).

7 An Optional Exercise

Explain in a concise manner, how Andrieu, and Doucet (1999) establish the acceptance probabilities for their reversible jump MCMC.

8 References and further reading:

1 Journal articles and technical reports:

- C. Andrieu, and A. Doucet (1999): Joint Bayesian Model Selection and Estimation of Noisy Sinusoids via Reversible Jump MCMC. *IEEE Transactions on Signal Prcessing*, 47, pp. 2667−2676.
- S.P. Brooks, P. Giudici, and G.O. Roberts (2003): Efficient construction of reversible jump MCMC proposal distributions. *Journal of the Royal Statistical Society, B*, 65, pp. 1−37.
- W.K. Hastings (1970): Monte Carlo sampling Methods Using Markov Chains and Their Applications. *Biometrika*, 57, pp. 97−109.
- U. Grenander, and M.I. Miller (1994): Representations of Knowledge in Complex Systems. *Journal of the Royal Statistical Society, B*, 56, pp. 549−603.
- R. Waagepetersen, and D. Sorensen (2001): A tutorial on reversible jump MCMC with a View toward Applications in QTL-Mapping. *International Statistical Review*, 69, pp.49−61.

2 Books:

- F.Gustafsson (2000): *Adaptive Filtering and Change Detection*, John Wiley & Sons, Chichester, Weinheim, New York, Brisbane, Singapore, Toronto.
- C.P. Robert (2001): *The Bayesian Choice. Second Edition*. Springer Verlag, New York.
- C.P. Robert and G. Casella (1999): *Monte Carlo Statistical Methods*. Springer Verlag, New York.

- D. Sorensen and D. Gianola (2001): *Likelihood, Bayesian, and MCMC Methods in Quantitative Genetics*. Statistics for Biology and Health. Springer Verlag, New York.