Reversible Jump MCMC (RJMCMC)

Most MCMC algorithms are constructed to sample from a target density with a fixed number of dimensions. In 1995, Peter Green [1] proposed a powerful new framework for the construction of "dimension jumping" algorithms known as **reversible jump MCMC (RJMCMC** algorithms. They are generalizations of the beloved Metropolis-Hastings algorithm. So...why might one want to dimension jump?

Example 1: Bayesian Model Selection

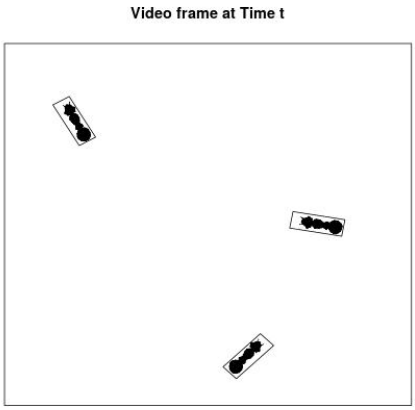
Given data, assume a countable collection of models $\{M_k : k \in I\}$ for some index set I. Further assume that model M_k has parameters $\vec{\theta}_k \in \mathbb{R}_{n_k}$ for some positive integer n_k.

The goal is to make inferences about the model and model parameters by drawing from a posterior distribution

$$\pi(k, \vec{\theta}_k | \text{ data })$$

Example 2: Multiple Target Tracking

Suppose that we are attempting to track fairly multiple target objects through a sequence of video frames with bounding rectangles.



For the fairly homogeneous looking targets shown, we might use a fixed size rectangle that can be described simply with a center (x, y) and an angle θ.

A target density of interest could then be

$$\pi(x_1^{(t)}, y_1^{(t)}, \theta_1^{(t)}, x_2^{(t)}, y_2^{(t)}, \theta_2^{(t)}, x_3^{(t)}, y_3^{(t)}, \theta_3^{(t)} | \text{ past tracking history}).$$

Now suppose that new objects can enter an leave the camera view over time. The target density of interest would now have the form

$$\pi(k^{(t)}, \vec{\theta}_k^{(t)} | \text{ past tracking history})$$

where k is the current number of objects being tracked and $\vec{\theta}_k$ is a $3k$-dimensional vector of parameters.

The Metropolis-Hastings (MH) Algorithm

The Metropolis-Hastings algorithm is a popular algorithm for drawing from a given distribution π on $\mathbb{R}^k$, which works by creating a Markov chain with π as its stationary or limiting distribution. The basic idea is, given a current state x, to potentially move to a state y that has been chosen from a candidate transition density $q(x, y)$ and to accept this move with a "corrective" probability that ultimately makes the overall transition density into some $p(x, y)$ which corresponds to the stationary distribution π. This means that

$$\pi(y) = \int \pi(x)p(x, y)\, dx.$$

Imagine running a chain according to transitions given by $p(x, y)$ out to time ∞. If $p(x, y)$ is "nice enough", the values seen at time ∞ will be settling down to draws from some well defined "limiting distribution". It is easy to show that such a limiting distribution is stationary. if $p(x, y)$ is "nice enough", there will only be one unique limiting distribution and so the limiting and stationary distributions are one and the same.

In order for a chain running according to a transition density $p(x, y)$ to have stationary distribution π, it is sufficient (but not necessary) for the **detailed balance** or **reversibility** condition to hold:

$$\pi(x)p(x, y) = \pi(y)p(y, x).$$

In this case, we can integrate both sides with respect to x to get

$$\int \pi(x)p(x, y)\, dx = \int \pi(y)p(y, x)\, dx = \pi(y) \int p(y, x)\, dx = \pi(y) \cdot 1 = \pi(y). \ \surd$$

The MH algorithm was derived out of trying to force reversibility.

Algorithm:

Let x be the current state of the Markov chain. (At time 0 this may be chosen arbitrarily.) Let $q(x, y)$ be a candidate transition density that lives on the same space as π in $\mathbb{R}^k$.

1. Generate a "candidate value" y according to $q(x, y)$.
2. Accept a move from x to y, and set $x = y$, with probability

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

3. Return to step 1.

After a "long time", stop the algorithm. The current state of the chain will be an approximate draw from π. (The state of the chain at time ∞ will be an exact draw from π.)

Note that the actual transitions of the MH chain take place according to a transition density

$$p(x, y) = q(y)\alpha(x, y), \qquad y \neq x$$

and with probability of remaining at the same point given by

$$p(x, \{x\}) = \int q(y)[1 - \alpha(x, y)]\, dy.$$

With this choice of α we have that π is stationary for p. While it is hard to directly verify that $\pi(y) = \int \pi(x)p(x, y)\, dx$, it is easy to verify that $\pi(x)p(x, y) = \pi(y)p(y, x)$.

Metropolis-Hastings Through a Different Lens

The yet to be described RJMCMC algorithm is a generalization of the MH algorithm where the acceptance probability $\alpha(x, y)$ involves a Jacobian term. This is often mistakenly thought to be a result of dimension jumping in RJMCMC when, in fact, it comes about just by writing down the MH algorithm in a different way.

In this Section it will be convenient for us to talk about moving from some x to some x' as opposed to some y. The detailed balance condition is

$$\pi(x)p(x, x') = \pi(x')p(x', x). \tag{1}$$

Fix any sets $A, A' \subseteq \mathbb{R}^k$. (1) clearly implies

$$\int_A \int_{A'} \pi(x)p(x, x')\, dx'\, dx = \int_A \int_{A'} \pi(x')p(x', x)\, dx'\, dx \tag{2}$$

as long as both integrals are well-defined.

It is possible that (2) holds but not (1). However, if (2) holds **for all** subsets $A, A' \subseteq \mathbb{R}^k$ the one can show that (1) must hold. [1]

If (2) holds for all subsets of $\mathbb{R}^k$ we say that the **integrated detailed balance condition** holds. We will rewrite the integrated detailed balance condition as

$$\boxed{\int_A \pi(x)P(x, A')\, dx = \int_{A'} \pi(x')P(x', A)\, dx', \qquad \forall\ A, A'} \tag{3}$$

where

$$P(x, B) := \int_B p(x, y)\, dy$$

will be referred to as a **transition kernel**.

Note that

$$P(x, A') = \int_{A'} q(x, x')\, \alpha(x, x')\, dx' + \left[\int q(x, x')[1 - \alpha(x, x')]\, dx'\right] \cdot I_{\{x \in A'\}}. \tag{4}$$

[1] Technically (2) must only hold for all Borel sets in $\mathbb{R}^k$ but this document is not all that technical!

That is to say, starting at x, we will be in the set A' after one iteration of the MH algorithm if we either

- propose a value $x' \in A'$ and accept the move, or
- propose a value anywhere, do not accept the move, thereby staying at x as long as x is already in A'. ($I_{x \in A'}$ is an indicator function.)

Let us denote the quantity in the square brackets in (4) as $R(x)$. The left-hand-side of (3) becomes

$$\int_A \pi(x) \left[\int_{A'} q(x,x')\, \alpha(x,x')\, dx' + R(x) \cdot I_{x \in A'} \right] dx = \int_A \int_{A'} \pi(x) q(x,x')\, \alpha(x,x') dx'\, dx + \int_{A \cap A'} \pi(x) R(x)\, dx$$

The right-hand-side of (3) becomes

$$\int_{A'} \pi(x') \left[\int_{A'} q(x',x)\, \alpha(x',x)\, dx + R(x') \cdot I_{x' \in A} \right] dx' = \int_A \int_{A'} \pi(x') q(x',x)\, \alpha(x',x) dx'\, dx + \int_{A \cap A'} \pi(x') R(x')\, dx'.$$

Note that those trailing "$R(x)$ integrals" are the same, thus, (3) becomes

$$\boxed{\int_A \int_{A'} \pi(x) q(x,x')\, \alpha(x,x') dx'\, dx = \int_A \int_{A'} \pi(x') q(x',x)\, \alpha(x',x) dx'\, dx} \tag{5}$$

Proposals Through Deterministic Functions

Note that the x' produced as a draw from $q(x,x')$ can be written as some deterministic function of x and some random variable/vector W where W has some probability density function x. For example, suppose we, when at x, propose a move to a value drawn from some normal distribution (with a given variance σ^2) centered at x. We might write

$$x' = x + w$$

where w is a realization of $W \sim N(0, \sigma^2)$.

Or, we might instead write

$$x' = x + \sigma w$$

where w is a realization of $W \sim N(0,1)$.

Or, we might instead write (Box-Muller!)

$$x' = x + \sigma \sqrt{-2 \ln w_1} \cos(2\pi w_2)$$

where $w = (w_1, w_2)$ is a realization of $W = (W_1, W_2)$ where $W_1, W_2 \overset{iid}{\sim} unif(0,1)$.

Or, we may write the proposal x' any of a number of other ways! For the purpose of illustration, we will move forward with the representation $x' = x + \sigma w$ for $W \sim N(0,1)$. Also, we will define a new random variable W', with realization w' so that our deterministic function update is dimension matching. That is, rather than writing

$$x' = h(x,w) \quad \text{(No!)}$$

for some $h : \mathbb{R}^2 \to \mathbb{R}$, we will write

$$(x', w') = h(x,w)$$

for some $h : \mathbb{R}^2 \to \mathbb{R}^2$.

(Outside of this particular normal example $x \in \mathbb{R}^k$ and $w \in \mathbb{R}^d$ for some positive integer d.)

How is W' defined? We wish to define it so that we can reverse the move using

$$(x, w) = h'(x', w')$$

with some function $h' : \mathbb{R}^2 \to \mathbb{R}^2$.

One possibility for this example is to define $W' = W$, in which case we have

$$\begin{array}{rcccl} (x', w') & = & h(x, w) & := & (x + \sigma w, w) \\ (x, w) & = & h'(x', w') & := & (x' - \sigma w', w'). \end{array}$$

Another possibility is to define $W' = -W$, in which case we have

$$\begin{array}{rcccl} (x', w') & = & h(x, w) & := & (x + \sigma w, -w) \\ (x, w) & = & h'(x', w') & := & (x' + \sigma w', -w'). \end{array}$$

In this case $h' \equiv h$. Also, in this case, W' is no longer the same as W but happens to have the same distribution by symmetry of the $N(0, 1)$ distribution.

(Outside of this particular normal example, W and W' may have different distributions with densities g and g', respectively.)

In summary for the general case,

- π lives on some subset of $\mathbb{R}^k$.
- $x, x' \in \mathbb{R}^k$.
- $w, w' \in \mathbb{R}^d$ for some d.
- w has some density g.
- x' and w' are defined through some function $h(x, w)$:

$$(x', w') = h(x, w)$$

- w' is further constrained by another function h' which is chosen so that the move can be reversed:

$$(x, w) = h'(x', w').$$

- When considering moving forward from (x, w) to (x', w'), x' and w' are both deterministic functions of x and w.
- If (x, w) is not known, the algorithm at state x' at time $t + 1$ is be defined in reverse to move to (produce) some (x, w) at time t. In this case w' needs to be drawn from some density g' which may or may not be equal to g.

Once h, h', g , and g' are defined, the candidate transition density may be defined as

$$q(x, x') = g(w)$$

for a suitable w. For our particular example,

$$q(x, x') = g(w) \text{ where } w \text{ is such that } x' = x + \sigma w$$

and $g(w) := N(w; 0, 1)$ is the $N(0, 1)$ pdf.

Back to Metropolis-Hastings

In our new notation, (5) may be rewritten

$$\boxed{\int_A \int_{B'} \pi(x) g(w)\, \alpha(x, x') dw\, dx = \int_B \int_{A'} \pi(x') g'(w')\, \alpha(x', x) dx'\, dw'} \tag{6}$$

where, for our running example, $x' = x + \sigma x$ and $B' = \{w : x + \sigma w \in A'\}$. The x on the right-hand side is an appropriate function of x' and w' and g' is defined appropriately.

Note that the right-hand-side of (6) can be rewritten with a substitution back to x and w using

$$dx'\, dw' = \left| \frac{\partial(x', w')}{\partial(x, w)} \right| dx\, dw \tag{7}$$

using the Jacobian

$$\frac{\partial(x', w')}{\partial(x, w)} := \begin{vmatrix} \frac{\partial x'_1}{\partial x_1} & \frac{\partial x'_1}{\partial x_2} & \cdots & \frac{\partial x'_1}{\partial x_k} & \frac{\partial x'_1}{\partial w_1} & \frac{\partial x'_1}{\partial w_2} & \cdot \frac{\partial x'_1}{\partial w_d} \\ \frac{\partial x'_2}{\partial x_1} & \frac{\partial x'_2}{\partial x_2} & \cdots & \frac{\partial x'_2}{\partial x_k} & \frac{\partial x'_2}{\partial w_1} & \frac{\partial x'_2}{\partial w_2} & \cdot \frac{\partial x'_2}{\partial w_d} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \frac{\partial x'_k}{\partial x_1} & \frac{\partial x'_k}{\partial x_2} & \cdots & \frac{\partial x'_k}{\partial x_k} & \frac{\partial x'_k}{\partial w_1} & \frac{\partial x'_k}{\partial w_2} & \cdot \frac{\partial x'_k}{\partial w_d} \\ \frac{\partial w'_1}{\partial x_1} & \frac{\partial w'_1}{\partial x_2} & \cdots & \frac{\partial w'_1}{\partial x_k} & \frac{\partial w'_1}{\partial w_1} & \frac{\partial w'_1}{\partial w_2} & \cdot \frac{\partial w'_1}{\partial w_d} \\ \frac{\partial w'_2}{\partial x_1} & \frac{\partial w'_2}{\partial x_2} & \cdots & \frac{\partial w'_2}{\partial x_k} & \frac{\partial w'_2}{\partial w_1} & \frac{\partial w'_2}{\partial w_2} & \cdot \frac{\partial w'_2}{\partial w_d} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \frac{\partial w'_d}{\partial x_1} & \frac{\partial w'_d}{\partial x_2} & \cdots & \frac{\partial w'_d}{\partial x_k} & \frac{\partial w'_d}{\partial w_1} & \frac{\partial w'_d}{\partial w_2} & \cdot \frac{\partial w'_d}{\partial w_d} \end{vmatrix}.$$

Substituting (7) into (6), it is easy to verify that the resulting equality will be satisfied by taking

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

This is the appropriate acceptance probability for the MH algorithm when viewed through the deterministic function update "lens". In order for everything to work, the derivatives in the Jacobian need to exist and the Jacobian needs to be non-zero.[2]

[2] h needs to be a smooth continuous bijective map with a smooth continuous inverse. We call such an h a **diffeomorphism**.

RJMCMC

As mentioned, the RJMCMC algorithm is a generalization of the MH algorithm with dimension jumping moves allowed. Suppose that we want to make a move from $x \in \mathbb{R}^k$ to $x' \in \mathbb{R}^{k'}$. We will again use random variables W and W' and a function h that takes (x, w) to (x', w') in a smooth, continuous bijection. To this end, we may need to "pad out" w or w' with auxiliary variables to final lengths of d and d', respectively, so that the following **dimension matching requirement holds**:

$$k + d = k' + d'.$$

As the RJMCMC algorithm really needs to be highly tailored to the specific target density we will describe it in the context of an example taken mostly from [2] with some liberties.

Running RJMCMC (Bayesian Model Selection) Example:

Suppose that we have some count data $X_1, X_2, \ldots, X_n$ that we believe is either from a Poisson distribution or a negative binomial distribution. In a Bayesian context, we may want to select the appropriate model based on some prior expert opinion about which model should be used and its parameters. Alternatively, if we reparameterize one or both models so that they have parameters with common interpretations, our Bayesian selves might want to return an estimate of parameter means, weighted by posterior model probabilities.

In particular, we will write Model 1:

$$X_1, X_2, \ldots, X_n \overset{iid}{\sim} Poisson(\lambda)$$

which has parameter $\lambda > 0$ and

$$P(X_i = x) = \frac{e^{-\lambda}\lambda^x}{x!} I_{\{0,1,2,\ldots\}}(x).$$

Model 2 will be a particular version of the negative binomial distribution with

$$P(X_i = x) = \left(\begin{array}{c} r + x - 1 \\ x \end{array} \right) p^r (1-p)^x I_{\{0,1,2,\ldots\}}(x),$$

which has parameters $0 \leq p \leq 1$ and $r \in \{1, 2, 3, \ldots\}$. Note that this can be generalized to non integer $r > 0$ by writing the pdf as

$$P(X_i = x) = \frac{\Gamma(r+x)}{x!\Gamma(r)} p^r (1-p)^x I_{\{0,1,2,\ldots\}}(x).$$

Note that Model 1 has mean λ and Model 2 has mean $r(1-p)/p$. In order to force a common parameter, which is only necessary if we want to give a model averaging result, we will take $p = r/(\lambda + r)$. Now, Model 1, parameterized by $\lambda > 0$ has mean λ and Model 2, parameterized by $\lambda > 0$ and $r > 0$ also has mean λ. Thus, we have a parameter with an interpretation that is the same for both models.

Let $K \in \{1, 2\}$ denote the model number. We have likelihoods given by $\pi(\vec{x}|k, \vec{\theta}_k)$. When $k = 1$, we are in the Poisson model with $\vec{\theta}_1 = \theta_1 := \lambda$. The likelihood is

$$\pi(\vec{x}|1, \vec{\theta}_1) = \pi(\vec{x}|1, \lambda) = \prod_{i=1}^{n} \frac{e^{-n\lambda}\lambda^{\sum x_i}}{x_i!} = \frac{e^{-n\theta_1}\theta_1^{\sum x_i}}{\prod(x_i!)}$$

When $k = 2$, we are in the negative binomial model with $\vec{\theta}_2 = (\theta_{21}, \theta_{22}) = (\lambda, r)$. The likelihood is

$$\begin{aligned} \pi(\vec{x}|2, \vec{\theta}_2) &= \pi(\vec{x}|2, \lambda, r) = \textstyle\prod_{i=1}^{n} \frac{\Gamma(r+x_i)}{x_i!\Gamma(r)} \left(\frac{r}{\lambda+r}\right)^r \left(\frac{\lambda}{\lambda+r}\right)^{x_i} \\ &= \left[\textstyle\prod_{i=1}^{n} \frac{\Gamma(r+x_i)}{x_i!}\right] \frac{1}{[\Gamma(r)]^n} \left(\frac{r}{\lambda+r}\right)^{rn} \left(\frac{\lambda}{\lambda+r}\right)^{\sum x_i}. \end{aligned}$$

Let $\pi(k)$ denote a prior distribution over models. For example, we may *a priori* believe that

$$\pi(k) = \begin{cases} 1/2 & , \quad k = 1 \\ 1/2 & , \quad k = 2 \end{cases}$$

Let $\pi(\vec{\theta}_k|k)$ denote a prior distribution for the parameters for Model k. For example, we may have a priori that

$$\pi(\vec{\theta}_1|1) = \pi(\vec{\lambda}|1) = \Gamma(\lambda; \alpha_\lambda, \beta_\lambda)$$

where $\Gamma(x; \alpha, \beta)$ denotes the gamma density with shape parameter α and inverse scale parameter β. For Model 2, if we make the simplifying assumption that λ and r are a priori independent, we might have

$$\pi(\vec{\theta}_2|2) = \pi(\vec{\lambda}, r|2) = \pi(\vec{\lambda}|2) \cdot \pi(r|2) = \Gamma(\lambda; \alpha_\lambda, \beta_\lambda) \cdot IG(r; \alpha_r, \beta_r)$$

where $IG(x; \alpha, \beta)$ denotes the inverse gamma density with parameters α and β so that $1/r \sim \Gamma(\alpha, \beta)$.

The target distribution for our RJMCMC simulation will be the posterior distribution:

$$\begin{aligned} \pi(k, \vec{\theta}_k|\vec{x}) &\propto \pi(\vec{x}|k, \vec{\theta}_k)\, \pi(\vec{\theta}_k|k)\, \pi(k) \\ &= \begin{cases} \pi(\vec{x}|1, \vec{\theta}_1)\, \pi(\vec{\theta}_1|1)\, \pi(1) & , \quad k = 1 \\ \pi(\vec{x}|2, \vec{\theta}_2)\, \pi(\vec{\theta}_2|2)\, \pi(2) & , \quad k = 2 \end{cases} \end{aligned}$$

The Moves

We will propose a move from $x = (k, \vec{\theta}_k)$ to $x' = (k', \vec{\theta}_{k'})$ in a way that assures that (5) holds. To this end, we will assume that the transition density $q(x, x') = q((k, \vec{\theta}_k), q)(k', \vec{\theta}_{k'})$ has the form

$$q((k, \vec{\theta}_k), q)(k', \vec{\theta}_{k'}) = q(\vec{\theta}_{k'}|\vec{\theta}_k, k, k') \cdot q(k'|k)$$

for user chosen $q(\vec{\theta}_{k'}|\vec{\theta}_k, k, k')$ and $q(k'|k)$. (We could make these dependent on the observed data $\vec{x}$ as well but this is complicated enough as is!)

For example, we might make equally likely moves to both models from both models with

$$q(k'|k) = \begin{cases} \frac{1}{2} & , \quad k' = 1 \\ \frac{1}{2} & , \quad k' = 2 \end{cases}$$

for any fixed value of $k \in \{1, 2\}$.

Transitions from the current state $x = (k, \vec{\theta}_k)$ will be proposed as follows.

1. Choose k' from $q(k'|k)$.
2. Choose $\vec{\theta}_{k'}$ from $q(\vec{\theta}_{k'}|\vec{\theta}_k, k, k')$.

Now, (5) has the form

$$\int_A \int_{A'} \pi(k, \vec{\theta}_k|\vec{x}) q(\vec{\theta}_{k'}|\vec{\theta}_k, k, k') q(k'|k)\, \alpha((k, \vec{\theta}_k), (k', \vec{\theta}'_k)) dx'\, dx \tag{8}$$

$$= \int_A \int_{A'} \pi(k', \vec{\theta}_{k'}|\vec{x}) q(\vec{\theta}_k|\vec{\theta}_{k'}, k, k') q(k|k')\, \alpha((k', \vec{\theta}'_k), (k, \vec{\theta}_k)) dx'\, dx$$

where A is any set of $(k, \vec{\theta}_k)$ points and A' is any set of $(k', \vec{\theta}_{k'})$ points. The notation is a little messy but $\int_A dx$ means to sum over all k and integrate over all $\vec{\theta}_k$ where $(k, \vec{\theta}_k) \in A$.

Model 1 to Model 1 or Model 2 to Model 2:

In the case where $k' \sim q(k'|k)$ is drawn to be equal to k (and the sets A and A' are appropriately restricted to reflect this) we can get equality of the integrals in (8) through equality of the integrands with traditional MH detailed balance:

$$\alpha((k', \vec{\theta}'_k), (k, \vec{\theta}_k)) = \min\left\{1,\ \frac{\pi(k', \vec{\theta}_{k'}|\vec{x}) q(\vec{\theta}_k|\vec{\theta}_{k'}, k, k') q(k|k')}{\pi(k, \vec{\theta}_k|\vec{x}) q(\vec{\theta}_{k'}|\vec{\theta}_k, k, k') q(k'|k)}\right\} \overset{k'=k}{=} \min\left\{1,\ \frac{\pi(k', \vec{\theta}_{k'}|\vec{x}) q(\vec{\theta}_k|\vec{\theta}_{k'}, k, k')}{\pi(k, \vec{\theta}_k|\vec{x}) q(\vec{\theta}_{k'}|\vec{\theta}_k, k, k')}\right\}$$

The ratio is precisely the old

$$\frac{\pi(y) q(y, x)}{\pi(x) q(x, y)} \overset{notation}{=} \frac{\pi(y) q(x|y)}{\pi(x) q(y|x)}$$

with π a conditional density given $\vec{x}$, x replaced by $(k, \vec{\theta}_k)$ and y essentially replaced by $(k', \vec{\theta}_{k'})$.

In the case where $k' \sim q(k'|k)$ is drawn to be not equal to k we can get equality of the integrals in (8) through a "nice" (diffeomorphism) mapping from $(\vec{\theta}_k, w)$ and a dimension matched $(\vec{\theta}_{k'}, w')$ change of variables which will result in matching integrands where one involves a Jacobian term.

Model 1 to Model 2:

In particular, let us first consider a move from Model 1 to Model 2 (i.e. $k = 1$ and $k' = 2$). That is, let us consider a move from $x = (1, \theta) = (1, \lambda)$ to $x' = (2, \theta') = (2, \lambda', r')$. The proposed transition is our choice and we will choose (for simplicity) to keep the λ for x' as it was in x (i.e. to set $\lambda' = \lambda$) and to propose an independent $r' \sim exp(rate = 1)$. Note that we can generate r by the inverse cdf method by generating $w \sim unif(0, 1)$ and setting $r = -\ln w$. (Here we are using the fact the inverse cdf for the exponential distribution with rate 1 is $F^{-1}(x) = -\ln(1 - x)$ and the fact that for $W \sim unif(0, 1)$, $1 - W \overset{d}{=} W$.)

Now

$$\theta' = (\lambda', r') = h(\lambda, w) := (\lambda, -\ln w)$$

where w is a realization of $W \sim unif(0, 1)$ with density $g(w) = I_{(0,1)}(w)$. Note that (λ, w) is two-dimensional and (λ', r') is two-dimensional. (Dimension matching check!)

The reverse move would be to set

$$(\lambda, w) = h'(\lambda', r') := (\lambda', e^{-r'}).$$

Note that this is reversal does not need for us to simulate any w' and is already dimension matched. This means that $g'(w')$ will not appear at all in the acceptance ratio.

In summary, if a move from Model 1 to Model 2 is proposed as a result of a draw from $q(k'|k)$, we proceed as follows.

- Draw $w \sim unif(0,1)$.
- Set $(\lambda', r') = (\lambda, -\ln w)$.
- Compute the Jacobian

$$\frac{\partial(\lambda', r')}{\partial(\lambda, w)} = \begin{vmatrix} \frac{\partial\lambda'}{\partial\lambda} & \frac{\partial\lambda'}{\partial w} \\ \frac{\partial r'}{\partial\lambda} & \frac{\partial r'}{\partial w} \end{vmatrix} = \begin{vmatrix} 1 & 0 \\ 0 & \frac{1}{w} \end{vmatrix} = \frac{1}{w}$$

- Accept the move from $x = (1, \lambda)$ to $x' = (2, \lambda', r')$ with probability

$$\alpha((1,\lambda),(2,\lambda',r')) = \min\left\{1,\ \frac{\pi(2,\lambda',r'|\vec{x})\, q(k|k')}{\pi(1,\lambda|\vec{x})\, g(w)\, q(k'|k)} \cdot \left|\frac{\partial(\lambda', r')}{\partial(\lambda, w)}\right|\right\}.$$

  Otherwise, stay at $x = (1, \lambda)$.

Model 2 to Model 1:

We know consider a move from Model 2 to Model 1 (i.e. $k = 2$ and $k' = 1$). That is, let us consider a move from $x = (2, \theta) = (2, \lambda, r)$ to $x' = (1, \theta') = (1, \lambda')$. The proposed transition is again our choice and we will choose the natural move to keep the λ for x' as it was in x (i.e. to set $\lambda' = \lambda$)

So, we should use

$$\lambda' = h(\lambda, r) = \lambda \quad \text{(No!)}$$

But, this is not dimension matching since $(\lambda, r) \in \mathbb{R}^2$ and $\lambda' \in \mathbb{R}$! We will fix this by simply reversing what we did in going from Model 1 to Model 2 and defining

$$(\lambda', w') = h'(\lambda, r) := (\lambda, e^{-r}).$$

The Jacobian is

$$\frac{\partial(\lambda', w')}{\partial(\lambda, r)} = \begin{vmatrix} \frac{\partial\lambda'}{\partial\lambda} & \frac{\partial\lambda'}{\partial r} \\ \frac{\partial w'}{\partial\lambda} & \frac{\partial w'}{\partial r} \end{vmatrix} = \begin{vmatrix} 1 & 0 \\ 0 & -e^{-r} \end{vmatrix} = -e^{-r}$$

The acceptance probability is

$$\alpha((2,\lambda,r),(1,\lambda')) = \min\left\{1,\ \frac{\pi(1,\lambda'|\vec{x})\, q(k|k')g(w')}{\pi(2,\lambda,r|\vec{x})\, q(k'|k)} \cdot \left|\frac{\partial(\lambda', r')}{\partial(\lambda, w)}\right|\right\}.$$

(Note: As a check, the acceptance probability ratio should be the reciprocal of the previously computed one after switching "primes" to "non-primes" and vice versa.)

References

[1] P.J. Green. Reversible jump Markov chain Monte Carlo computation and bayesian model determination. *Biometrika*, 82:771–732, 1995.

[2] D.I. Hastie and P.J. Green. Model choice using reversible jump Markov chain Monte Carlo. *Statistica Neerlandica*, 66:309–338, 2012.