note0401

Daniel Sorensen

November 10, 2024

Appendix: A closer look at the McMC machinery

The appendix provides a brief description of the Metropolis-Hastings algorithm in its standard and general form. The intention is to provide an intuition for the rationale behind the form of the algorithm. The material may seem a little daunting at first sight, particularly due to the nature of the ideas and the notation than the mathematics. In fact, the appendix uses nuts and bolts mathematics and basic theory of transformation of random variables. Much of what is written below is taken from Waagepetersen and Sorensen (2001), where more details and an example can be found. A benchmark paper for the general Metropolis-Hastings algorithm is Green (1995); the paper is technical and its detailed understanding requires command of measure theory. An accessible overview of discrete and continuous Markov chains can be found in chapters 10 and 11 of Sorensen and Gianola (2002).

The standard Metropolis-Hastings ratio

Consider first a finite state discrete Markov chain with state space S. The Markov chain is defined by a sequence of discrete random variables X_i, $i = 1, 2, \ldots$ each of which can take one of N values in the finite set $S = \{0, 1, \ldots, N-1\}$. The subscript i in X_i can be interpreted as stages or time periods, and the initial stage is $i = 0$. These random variables X_i satisfy the Markov property

$$\Pr(X_i|X_{i-1}, X_{i-2}, \ldots, X_0) = \Pr(X_i|X_{i-1}). \tag{4.92}$$

Stationarity

A probability density π is *stationary* or *invariant*, if $X_i \sim \pi$ implies $X_{i+1} \sim \pi$, $i \geq 0$. Notationally x and x' often will be used to represent values of the random variable X at two time periods or stages (not to be confused with the same symbol "$'$" used elsewhere to denote the transpose of a matrix or vector). A definition of stationarity is

$$\sum_x \pi(x) \Pr(X_{i+1} = x'|X_i = x) = \pi(x'), \quad x, x' \in S. \tag{4.93}$$

The idea behind stationarity is as follows. Consider the relationship

$$\Pr(X_{i+1} = x'|X_0 = x_0) = \sum\nolimits_x \Pr(X_{i+1} = x'|X_i = x)\Pr(X_i = x|X_0 = x_0). \tag{4.94}$$

If the chain converges to a stationary distribution π that is independent of the starting point x_0, then

$$\begin{aligned} \Pr(X_{i+1} = x'|X_0 = x_0) &= \pi(x'), \\ \Pr(X_i = x|X_0 = x_0) &= \pi(x) \end{aligned}$$

and (4.94) reduces to (4.93).

The Metropolis-Hastings (MH) algorithm is a recipe for constructing a Markov chain that has π as its *stationary or invariant distribution*. The goal is to obtain draws from π which could be a posterior distribution. Since doing this directly may be very complicated or impossible, a draw from a *proposal distribution* (denoted here as $q(\cdot|x)$ that may depend on x) is taken instead and because this is not the same distribution as π, the drawn value is accepted in a stochastic manner by means of an *acceptance probability*. This acceptance probability is derived in a way ensuring that the sequence $X_0, X_1, \ldots,$ is an approximate Monte Carlo sample from the stationary distribution π. In order for this to hold, the Markov chain must be aperiodic and irreducible and hence, ergodic (an aperiodic chain does not return to the same state at regular time intervals; in an irreducible chain, every state is reachable from every other state in a finite number of transitions). An ergodic chain is one that converges to the stationary distribution, regardless of the starting value. This Markov chain can be used for Monte Carlo estimation of various expectations with respect to the stationary distribution. Before showing how the algorithm works, I define the concept of detailed balance on which the MH algorithm builds.

Detailed balance

Let $T(x'|x) = \Pr(X_{i+1} = x'|X_i = x)$ be the conditional probability that $X_{i+1} = x'$, given that $X_i = x$. This conditional pmf is known as a *transition probability*, that has the standard property

$$\sum\nolimits_{x'} T(x'|x) = 1. \tag{4.95}$$

In the case of finite state discrete Markov chains, $T(x'|x)$ is an element in an $N \times N$ stochastic matrix (transition probability matrix) and (4.95) indicates that the sum of the elements of the rows of the stochastic matrix adds to one. Then π satisfies detailed balance with respect to T, if

$$\pi(x)T(x'|x) = \pi(x')T(x|x'). \tag{4.96}$$

The left hand side is equal to the right hand side with x and x' reversed. Note that when $x = x'$, (4.96) holds trivially because

$$\pi(x')T(x'|x') = \pi(x')T(x'|x'). \tag{4.97}$$

The detailed balance equation (4.96) can also be written as

$$\Pr(X_i = x)\Pr(X_{i+1} = x'|X_i = x) = \Pr(X_i = x')\Pr(X_{i+1} = x|X_i = x')$$

or

$$\Pr(X_i = x, X_{i+1} = x') = \Pr(X_i = x', X_{i+1} = x),$$

provided $X \sim \pi$.

A Markov chain that has stationary distribution π that satisfies detailed balance with respect to T is said to be a reversible Markov chain.

An intuition for expression (4.96) is that the total probability mass in the move from x to x' is equal to that of the reverse move, from x' to x. In the left hand side, the probability mass at x is $\pi(x)$ and only a proportion $T(x'|x)$ moves to the right hand side. The total probability mass in the move from x to x' is the product of these two quantities. Likewise, the probability mass on the right hand side at $X_i = x'$ is $\pi(x')$ that may be different from $\pi(x)$. A proportion of $\pi(x')$ equal to $T(x|x')$ moves from x' to x and the total probability mass is again the product of these two quantities. The total probability mass for this pair of states (x, x') on both sides is the same. This holds for all possible pairs x and x' that belong in S.

An important consequence of imposing the strong condition of detailed balance (4.96), is that an ergodic Markov chain that is reversible has π as its stationary distribution when it converges. To show this, start with (4.96) and sum over x on both sides (if (4.96) holds, the equality still holds if we sum both sides over x):

$$\begin{aligned}\sum_x \pi(x)T(x'|x) &= \sum_x \pi(x')T(x|x') \\ &= \pi(x')\sum_x T(x|x') \\ &= \pi(x'), \end{aligned} \tag{4.98}$$

that is the definition of a stationary distribution. In the case of a Markov chain in continuous space with transition kernel p, the equivalent to (4.98) is obtained by integrating both sides with respect to x:

$$\begin{aligned}\int \pi(x)p(x'|x)dx &= \int \pi(x')p(x|x')dx \\ &= \pi(x')\int p(x|x')dx \\ &= \pi(x')\end{aligned}$$

that again is the definition of a stationary distribution. This means that if (X_i, X_{i+1}) has stationary distribution π, then the time-reversed subchain (X_{i+1}, X_i) has the same stationary distribution, whenever X_i has density π.

The acceptance probability of the MH algorithm is derived assuming that the Markov chain satisfies detailed balance. If the Markov chain is ergodic, when it converges, detailed balance guarantees that π is the stationary distribution.

The acceptance probability of the Metropolis-Hastings algorithm

Consider a move from x to x'. In the MH algorithm a proposed value y for x' is drawn from the proposal distribution $q(y|x)$. If the proposed value is accepted, $x' = y$ and if its is rejected, $x' = x$, the value at the previous stage of the chain. There are two ways in which the state in the next stage is equal to x'. One is to draw the proposal y and to accept it with probability $a(y|x)$. The other way in which the state at the next stage can take the value x', is to reject the proposal, to set $x' = x$ but X was already equal to x'. Let

$$s(x) = \Pr(y \text{ rejected } |X_i = x) \tag{4.99}$$

Then

$$\begin{aligned} T(y|x) &= \Pr(X_{i+1} = y|X_i = x) \\ &= q(y|x)a(y|x) + s(x)I(x = y). \end{aligned}$$

The left hand side of (4.96) can now be written as

$$\pi(x)T(y|x) = \pi(x)q(y|x)a(y|x) + \pi(x)s(x)I(x = y), \tag{4.100}$$

and by symmetry the right hand side is equal to

$$\pi(y)T(x|y) = \pi(y)q(x|y)a(x|y) + \pi(y)s(y)I(y = x). \tag{4.101}$$

The second terms in the right hand side of (4.100) and (4.101) are equal, both in the case when $x \neq y$ in which case they are zero because the indicator function is zero, or trivially when $x = y$. Therefore detailed balance is satisfied if

$$\pi(x)q(y|x)a(y|x) = \pi(y)q(x|y)a(x|y). \tag{4.102}$$

Then,

$$\frac{a(y|x)}{a(x|y)} = \frac{\pi(y)q(x|y)}{\pi(x)q(y|x)}. \tag{4.103}$$

Relationship (4.103) is satisfied if $a(y|x) = \pi(y)q(x|y)/b(y|x)$, for some $b(y|x) \geq \pi(y)q(x|y)$, to ensure that the acceptance probability $a(y|x) \leq 1$ (a similar argument holds for $a(x|y)$ due to symmetry). A valid $b(y|x)$ is

$$b(y|x) = \max(\pi(y)q(x|y), \pi(x)q(y|x)) \tag{4.104}$$

but other choices are possible and these have an impact on the properties of the Markov chain. The subject is rather technical and is discussed by Hastings (1970) and Peskun (1973), where it is shown that (4.104) leads to a Markov chain with the largest possible acceptance probabilities resulting in minimum asymptotic variances of moment estimates. This choice translates into the most commonly cited expression for the Metropolis-Hastings acceptance probability, given by

$$a(y|x) = \min\left(1, \frac{\pi(y)q(x|y)}{\pi(x)q(y|x)}\right), \tag{4.105}$$

that gives rise to the following algorithm:

- Initialise setting $X = x$, so that $\pi(x) > 0$
- Choose the proposal distribution q

  Given current state $X_i = x$, go through the loop
- Draw y from the proposal $q(\cdot|x)$
- Draw u from $Un(0, 1)$
- If $u < a(y|x)$, accept y and set $X_{i+1} = y = x'$, otherwise $X_{i+1} = x$

The original idea of using Markov chain simulation of probability distributions is often attributed to Metropolis et al. (1953).

The general Metropolis-Hastings ratio

In the standard Metropolis-Hastings algorithm, the next state x' is obtained by drawing the candidate y from q. A more general mechanism to generate the move to x', is to construct a proposal y' by applying a deterministic mapping to x and to a random component u, that has density $q_U(u|x)$, which may depend on x (Green, 1995). The proposal y' is

$$y' = g_1(x, u).$$

If the proposal is accepted, $x' = y'$. To ensure dimension matching in the move from (x, u) to x', if x is of size n_x and u of size n_u, it may be necessary to include a random variable u' of size $n_{u'}$ with density $q_{U'}(u'|x')$ so that

$$n_x + n_u = n_{x'} + n_{u'}. \tag{4.106}$$

Then the vectors of Markov chain states and proposal random variables (x, u) and (x', u') are of equal dimension and the densities $\pi(x)q_U(u|x)$ in the move from (x, u) to (x', u'), and $\pi(x')q_{U'}(u'|x')$ in the move from (x', u') to (x, u) are joint densities on spaces of equal dimension. The mapping is then

$$(x', u') = g(x, u) = (g_1(x, u), g_2(x, u)) \tag{4.107}$$

and the move in the opposite direction is

$$(x, u) = g^{-1}(x', u') = \big(g_1^{-1}(x', u'), g_2^{-1}(x', u')\big). \tag{4.108}$$

A necessary condition for the existence of the one-to-one mapping is that (4.106) is satisfied.

An **important detail** that becomes obvious in the derivation (more below), is that there is a constraint in the form of the deterministic function. The constraint is that

$$g = g^{-1}. \tag{4.109}$$

Therefore, we need

$$\begin{aligned}(x', u') &= g(x, u) = (g_1(x, u), g_2(x, u)), \\ (x, u) &= g^{-1}(x', u') = g(x', u') = (g_1(x', u'), g_2(x', u')).\end{aligned} \tag{4.110}$$

The general Metropolis-Hastings acceptance ratio takes the form

$$a(x'|x) = \min\left\{1, \frac{\pi(x')q_{U'}(u'|x')}{\pi(x)q_U(u|x)}|J|\right\} \tag{4.111}$$

where $|J| = \left|\det \frac{\partial g(x,u)}{\partial(x,u)}\right|$ is the absolute value of the Jacobian of the transformation g. A detailed derivation is given below.

Stationarity

The Markov chain in continuous space is specified in terms of the distribution for the initial state X_0 and the transition kernel $P(\cdot, \cdot)$ which specifies the conditional distribution of X_{t+1} given the previous state X_t. If the current state is $X_t = x$, then the probability that X_{t+1} is in a subset $B \subseteq \mathbb{R}^d$ is given by

$$P(x, B) = \Pr(X_{t+1} \in B|X_t = x). \tag{4.112}$$

Assume that π is a complex target distribution for a stochastic vector Z of dimension $d \geq 1$. Since expectations with respect to π cannot be evaluated analytically or by using techniques for numerical integration, a Markov chain X_i, $i = 1, 2, \ldots$, is constructed whose stationary distribution is π. If the chain is ergodic then it can be used for Monte Carlo estimation of expectations $\mathrm{E}(h(Z))$ for any function h, with respect to the invariant density π. That is

$$\mathrm{E}(h(Z)) = \int h(z)\pi(z)dz \approx \frac{1}{n}\sum\nolimits_{i=1}^{n} h(X_i) \tag{4.113}$$

as n tends to infinity. Thus $\mathrm{E}(h(Z))$ can be approximated by the sample average for large n, the length of the chain. The autocorrelation among the draws of the chain implies that the size of the Markov chain must be larger than when the draws are independent, in order to achieve a given level of accuracy.

For a continuous space Markov chain, the definition of stationarity is

$$\int P(x, B)\pi(x)dx = \int_B \pi(x)dx = \Pr(X \in B) \tag{4.114}$$

The distribution π is invariant (or stationary) for the Markov chain, if the transition kernel $P(\cdot, \cdot)$ of the Markov chain preserves π, so that $X_t \sim \pi$ implies $X_{t+1} \sim \pi$. In order to verify that π is the invariant density using (4.114) is an infeasible task, since this involves integration with respect to π. The difficulty of doing this was the reason for using McMC in the first place. However, choosing a kernel that imposes the stronger condition of reversibility with respect to π is sufficient to guarantee that π is the invariant density of the ergodic Markov chain.

Reversibility

For a continuous state space Markov chain, the reversibility condition

$$P_{t,t+1}(X_t \in A, X_{t+1} \in B) = P_{t,t+1}(X_t \in B, X_{t+1} \in A) \tag{4.115}$$

requires that the equilibrium probability that the state of the chain belongs to A and moves to B must be the same with A and B reversed, for subsets $A, B \subseteq \mathbb{R}^d$. In other words, expression (4.115) states that the joint probability that $X_t \in A$ and $X_{t+1} \in B$ (left hand side) is the same as the joint probability that $X_t \in B$ and $X_{t+1} \in A$ (right hand side). The left hand side will be referred to as the move from x to x' and the left hand side as the opposite move.

The reversibility condition (4.115) can also be written in terms of the transition kernel

$$\int_A P(x, B)\pi(x)dx = \int_B P(x, A)\pi(x)dx. \tag{4.116}$$

Reversibility (4.116) implies (4.114) by taking $A = \mathbb{R}^d$. Then $P(x, A) = 1$ and (4.116) reduces to

$$\int P(x, B)\pi(x)dx = \int_B \pi(x)dx$$

which is equal to (4.114). Therefore, an ergodic Markov chain that satisfies (4.116) has stationary distribution π.

The acceptance probability for a general Metropolis-Hastings algorithm

In a general setting, instead of generating x' from $q(\cdot|x)$ as is practised in the standard Metropolis-Hastings algorithm, x' can be defined in terms of a stochastic component $u \sim q(u|x)$ and a deterministic mapping g. As explained in connection with (4.107) and (4.108), in the move from x to x' the mapping is

$$(x', u') = g(x, u) = (g_1(x, u), g_2(x, u)) \tag{4.117}$$

and in the reverse move

$$(x, u) = g^{-1}(x', u') = (g_1(x', u'), g_2(x', u')). \tag{4.118}$$

The transition kernel in the move from $X_t \in A$ to $X_{t+1} \in B$ (move from x to x')

$$P(x, B) = \Pr(X_{t+1} \in B | X_t = x) = \int I(x' \in B)p(x'|x)dx'$$

is now constructed in three steps. The Metropolis-Hastings protocol requires first drawing u from $q(\cdot|x)$, secondly, constructing the proposal $y' = g_1(x, u)$ and thirdly accepting it with probability $a(g_1(x, u)|x)$. Define

$$S(x) = \Pr(y' \text{ is rejected } | X_t = x).$$

Then

$$\Pr(X_{t+1} \in B | X_t = x) = \int I(g_1(x,u) \in B) q(u|x) a(g_1(x,u)|x) du \\ + S(x) I(x \in B), \tag{4.119}$$

where the second term in the right hand side specifies the probability of rejecting y' but the current state x already belongs to the subset B. The left hand side of (4.115) takes the form

$$\begin{aligned} & \int_A \Pr(X_{t+1} \in B | X_t = x) \pi(x) dx \\ = & \int \int I(x \in A, g_1(x,u) \in B) \pi(x) q(u|x) a(g_1(x,u)|x) dx du \\ & + \int S(x) I(x \in A \cap B) \pi(x) dx. \end{aligned} \tag{4.120}$$

(The second term in the right hand side is equal to $\int_A S(x) I(x \in B) \pi(x) dx$).

A transition from $X_t \in B$ to $X_{t+1} \in A$ (move from x' to x) is accomplished by first drawing u' from $q(\cdot|x')$, secondly, constructing the proposal $y = g_1(x', u')$ and thirdly accepting it with probability $a(g_1(x', u')|x')$. The right hand side of (4.115) takes the form

$$\begin{aligned} & \int \int I(x' \in B, g_1(x', u') \in A) \pi(x') q(u'|x') a(g_1(x', u')|x') dx' du' \\ & + \int S(x') I(x' \in A \cap B) \pi(x') dx'. \end{aligned} \tag{4.121}$$

The second terms in (4.120) and (4.121) are equal (x and x' are dummy variables of integration) and therefore a sufficient condition for (4.115) to hold is

$$\begin{aligned} & \int \int I(x \in A, g_1(x,u) \in B) \pi(x) q(u|x) a(g_1(x,u)|x) dx du \\ = & \int \int I(x' \in B, g_1(x', u') \in A) \pi(x') q(u'|x') a(g_1(x', u')|x') dx' du'. \end{aligned} \tag{4.122}$$

The final step is to find a way to equalise the indicator functions of both sides of equation (4.122). This is achieved performing the change of variable

$$(x, u) = g(x', u') = (g_1(x', u'), g_2(x', u')) \tag{4.123}$$

and making use of

$$g^{-1}(x', u') = g(x', u'). \tag{4.124}$$

Substituting in the argument of the indicator function of the right hand side of (4.122), both indicator functions are equalised. It is at this point of the derivation that the constraint

(4.109) or (4.124) becomes relevant (see also (4.110)). Using this transformation and setting $dx'du' = |J|dxdu$, the right hand side of (4.122) takes the form

$$\iint I(g_1(x,u) \in B, x \in A)\pi(g_1(x,u))q(g_2(x,u)|g_1(x,u))a(x|g_1(x,u))|J|dxdu, \qquad (4.125)$$

where

$$J = \det\begin{bmatrix} \frac{\partial g_1(x,u)}{\partial x} & \frac{\partial g_2(x,u)}{\partial x} \\ \frac{\partial g_1(x,u)}{\partial u} & \frac{\partial g_2(x,u)}{\partial u} \end{bmatrix}.$$

Examination of the left hand side of (4.122) and of (4.125) shows that the reversibility condition (4.115) is satisfied if

$$\pi(x)q(u|x)a(g_1(x,u)|x) = \pi(g_1(x,u))q(g_2(x,u)|g_1(x,u))a(x|g_1(x,u))|J|.$$

The same argument leading to (4.105) indicates that a valid choice for $a(x'|x)$ is

$$a(x'|x) = \min\left[1, \frac{\pi(x')q(u'|x')}{\pi(x)q(u|x)}|J|\right]. \qquad (4.126)$$

A toy example

The model is $[y|\mu, \lambda] \sim N(\mu, \lambda)$. Assume μ is known and a Metropolis-Hastings algorithm is constructed to update the variance λ.

Strategy 1

This is accomplished generating $u \sim Un(a, b)$ and letting $\lambda' = \lambda u$. The current state of the Markov chain is $(z, u) = (\lambda, u)$ and the move is to

$$\begin{aligned} \left(z', u'\right) &= g(\lambda, u) \\ &= ((\lambda u), 1/u) \end{aligned}$$

where $u = 1/u'$. The inverse function that makes the move in the opposite direction possible is

$$\begin{aligned} (z, u) &= g^{-1}\left(z', u'\right) \\ &= g(\lambda', u') \\ &= ((\lambda' u'), 1/u') \end{aligned}$$

where $u' = 1/u$ is also generated from $Un(a, b)$. The Jacobian of the transformation $g(\lambda, u) = (\lambda u, 1/u)$ is

$$\begin{aligned} J &= \left|\det\left[\frac{\partial g(\lambda, u)}{\partial(\lambda, u)}\right]\right| \\ &= \left|\det\begin{bmatrix} u & \lambda \\ 0 & -u^{-2} \end{bmatrix}\right| = u^{-1}. \end{aligned}$$

Since u and u' are drawn from $Un(a,b)$, $1/(b-a)$ cancels in the ratio $q_{U'}(u'|z')/q_U(u|z)$ in (4.111) and the Metropolis-Hastings acceptance probability is

$$\min\left\{1, \frac{p(\mu, \lambda'|y)}{p(\mu, \lambda|y)} u^{-1}\right\}, \qquad u \in (a,b). \tag{4.127}$$

NOTE

The above strategy satisfies the constraint (4.109) which in the example takes the form

$$\begin{aligned} (z', u') &= \left((zu), \frac{1}{u}\right) \\ (z, u) &= \left((z'u'), \frac{1}{u'}\right) \end{aligned}$$

with $u = 1/u'$. This is so because (recall, $g_1(a,b) = a \times b$, is a function that multiplies its arguments)

$$\begin{aligned} z' &= g_1(z, u) \\ &= g_1(g_1(z', u'), u) \\ &= g_1(z', u') \times u \\ &= (z' \times u') \times u \end{aligned}$$

and therefore $u = 1/u'$.

Strategy 2

An alternative way of arriving at (4.127) is to use as the Metropolis-Hastings ratio

$$\frac{p(\mu, \lambda u|y)}{p(\mu, \lambda|y)} \frac{p(\lambda|\lambda')}{p(\lambda'|\lambda)}. \tag{4.128}$$

This is the standard form of the Metropolis-Hastings ratio, using the proposal for the parameter rather than the auxiliary variables (u, u'). Then with $\lambda' = u\lambda$,

$$p(\lambda'|\lambda) = q_U(u)\left|\frac{du}{d\lambda'}\right| = \frac{1}{b-a}\frac{1}{\lambda}, \qquad \lambda' \in (\lambda a, \lambda b).$$

By symmetry,

$$p(\lambda|\lambda') = \frac{1}{b-a}\frac{1}{\lambda'}, \qquad \lambda \in (\lambda' a, \lambda b).$$

The bounds $\lambda \in (\lambda' a, \lambda' b)$ imply

$$\begin{aligned} \lambda u' a &< \lambda < \lambda u' a, \\ a &< \frac{1}{u'} < b. \end{aligned}$$

The Metropolis-Hastings ratio is now

$$\begin{aligned}
\frac{p(\mu, \lambda u|y)}{p(\mu, \lambda|y)} \frac{p(\lambda|\lambda')}{p(\lambda'|\lambda)} &= \frac{p(\mu, \lambda u|y)}{p(\mu, \lambda|y)} \frac{\lambda}{\lambda'} \\
&= \frac{p(\mu, \lambda u|y)}{p(\mu, \lambda|y)} \frac{1}{u}, \qquad u \in (a, b).
\end{aligned} \tag{4.129}$$

and the resulting acceptance probability is

$$\min\left\{1, \frac{p(\mu, \lambda'|y)}{p(\mu, \lambda|y)} \frac{1}{u}\right\}, \qquad u \in (a, b). \tag{4.130}$$

Strategy 3

The third strategy consists of updating the variance using a random walk proposal density on the logvariance. That is

$$\ln \lambda' \sim N(\ln \lambda, k),$$

a normal distribution with mean equal to the natural logarithm of the previous realisation of λ and variance given by k, a user-tuned parameter.

As a reminder, if X has density $p(x) = N(m, k)$ and $Y = f(X) = \exp(X)$, such that the inverse function f^{-1} exists and results in $X = f^{-1}(Y) = \ln Y$, then the Jacobian of the transformation from X to Y is $1/y$ and $p(y) = p(f^{-1}(y))\frac{1}{y}$. In this particular case, in the move from λ to λ' we have $X = \ln \lambda'$; $Y = \exp(\ln \lambda') = \lambda'$; $f^{-1}(Y) = \ln \lambda'$. Therefore, if

$$q(\ln \lambda' | \ln \lambda, k) = N(\ln \lambda, k),$$

then

$$\begin{aligned}
q(\lambda' | \ln \lambda, k) &= q(\ln \lambda' | \ln \lambda, k) \frac{1}{\lambda'} \\
&= N(\ln \lambda, k) \frac{1}{\lambda'},
\end{aligned} \tag{4.131}$$

which is the density of the lognormal distribution with parameters $(\ln \lambda, k)$. In these expressions, the variance of the normal distribution k is a user-tuned parameter. Then (4.131) is the proposal density evaluated at λ' and by symmetry,

$$q(\lambda | \ln \lambda, k) = q(\ln \lambda | \ln \lambda', k) \frac{1}{\lambda}.$$

Since $q(\ln \lambda' | \ln \lambda, k) = q(\ln \lambda | \ln \lambda', k)$ the Metropolis-Hastings ratio is

$$\frac{p(\mu, \lambda'|y)}{p(\mu, \lambda|y)} \frac{q(\lambda | \ln \lambda, k)}{q(\lambda' | \ln \lambda, k)} = \frac{p(\mu, \lambda'|y)}{p(\mu, \lambda|y)} \frac{\lambda'}{\lambda},$$

different from (4.129).

References

Green, P. (1995). Reversible jump MCMC computation and Bayesian model determination. *Biometrika 82*, 711–732.

Hastings, W. K. (1970). Monte Carlo sampling methods using Markov chains and their application. *Biometrika 57*, 97–109.

Metropolis, N., A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller (1953). Equations of state calculations by fast computing machines. *Journal of Chemical Physics 21*, 1087–1092.

Peskun, P. H. (1973). Optimum Monte Carlo sampling using Markov chains. *Biometrika 60*, 607–612.

Sorensen, D. and D. Gianola (2002). *Likelihood, Bayesian, and MCMC Methods in Quantitative Genetics*. Springer-Verlag. 740 pp., Reprinted with corrections, 2006.

Waagepetersen, R. and D. Sorensen (2001). A tutorial on reversible jump MCMC with a view towards applications in QTL-mapping. *International Statistical Review 69*, 49–61.