This means that a candidate θ' that has a higher value of the target density than target density of the current value θ will always be accepted. The chain will always move "uphill." On the other hand, a candidate with a lower target value will only be accepted with a probability equal to the proportion of target density value to current density value. There is a certain probability that the chain will move "downhill." This allows a chain with a random-walk candidate density to move around the whole parameter space over time. However a chain with a random-walk candidate density will generally have many accepted candidates, but most of the moves will be a short distance. Thus, it might take a long time to move around the whole parameter space.

Example 8 *Suppose we have a unscaled target density given by*

$$g(\theta|y) = .8 \times e^{-\frac{1}{2}\theta^2} + .2 \times \frac{1}{2} e^{-\frac{1}{2\times 2^2}(\theta-3)^2} .$$

This is a mixture of a normal$(0, 1^2)$ and a normal$(3, 2^2)$. However, we only need to know the unscaled target since multiplying by a constant would multiply both the numerator and denominator by the constant which would cancel out. Let us use the normal candidate density with variance $\sigma^2 = 1$ centered around the current value as our random-walk candidate density distribution. Its shape is given by

$$q(\theta, \theta') = e^{-\frac{1}{2}(\theta'-\theta)^2} .$$

Let the starting value be $\theta = 2$. Figure 6.1 shows six consecutive draws from a random-walk Metropolis-Hastings chain together with the unscaled target and the candidate density. In a Metropolis-Hastings chain with a random-walk candidate density, the candidate density centered around the starting value, so it changes every time a candidate is accepted. Since the random-walk candidate density is symmetric about the current value, the acceptance probability

$$\begin{aligned} \alpha &= \min\left(\frac{g(\theta'|y)}{g(\theta|y)} \times \frac{q(\theta', \theta)}{q(\theta, \theta')}, 1\right) \\ &= \min\left(\frac{g(\theta'|y)}{g(\theta|y)}, 1\right) . \end{aligned}$$

We see it only depends on the target density. It always moves "uphill," and moves "downhill" with some probability. The candidate is accepted if a random draw from a uniform$(0, 1)$ is less than α. Table 6.1 gives a summary of the first six draws of this chain. Figure 6.2 shows the traceplot and a histogram for 1000 draws from the Metropolis-Hastings chain using a random-walk candidate density with standard deviation equal to 1. We see that the chain is moving through the space satisfactorily. We note that when it occasionally goes into the tail region, it does not stay there for long because the tendency to go "uphill" moves it back towards the central region that has higher values of the probability density. We see that the histogram has a way to go before it is close to the true posterior. Figure 6.3 shows the histograms for 5000 and 20000 draws from the Metropolis-Hastings chain. We see the chain is getting closer to the true posterior as the number of draws increases.

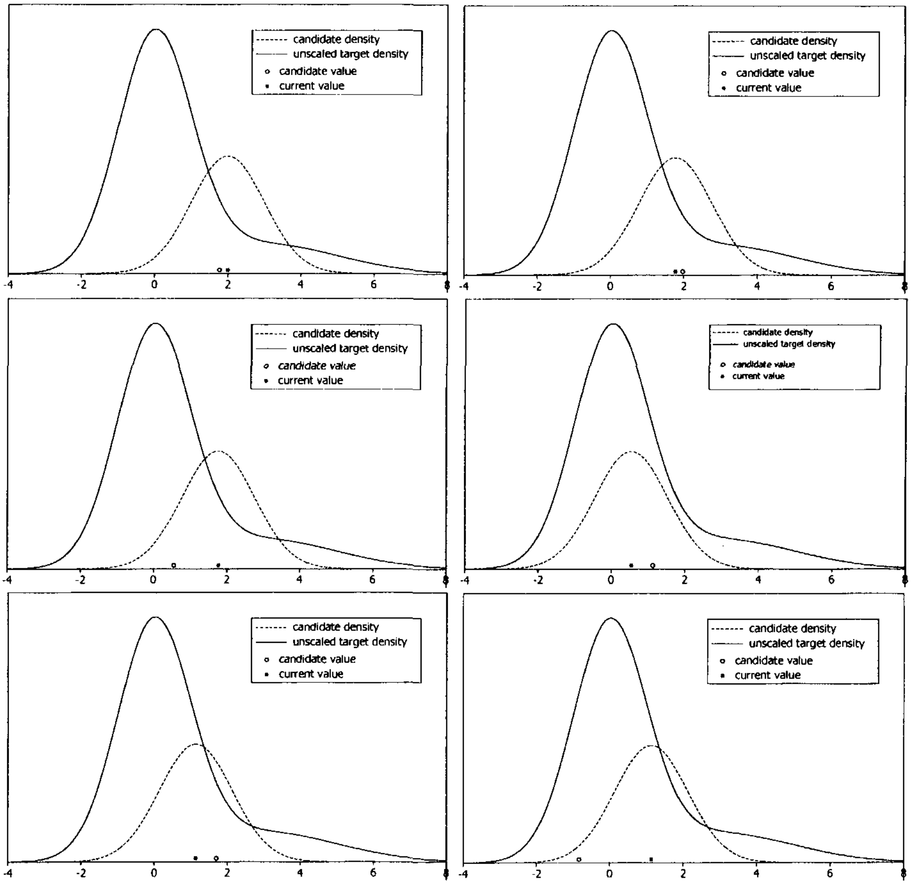

Figure 6.1 Six consecutive draws from a Metropolis-Hastings chain with a random-walk candidate density. Note: the candidate density is centered around the current value.

With random-walk candidate density function, the Markov chain will accept a large proportion of candidates. However, the accepted candidate will be close to the previous current value. The Markov chain will move through the state space (parameter space), however it may not be very fast.

Single Parameter with an Independent Candidate Density

Hastings (1970) introduced Markov chains with candidate generating density that did not depend on the current value of the chain. These are called *independent* candidate distribution

$$q(\theta, \theta') = q_2(\theta')$$

Table 6.1 Summary of first six draws of the chain using the random-walk candidate density

Draw	Current value	Candidate	α	u	Accept
1	2.000	1.767	1.000	.773	yes
2	1.767	1.975	.804	.933	no
3	1.767	.547	1.000	.720	yes
4	.547	1.134	.659	.240	yes
5	1.134	1.704	.553	.633	no
6	1.134	-.836	1.000	.748	yes

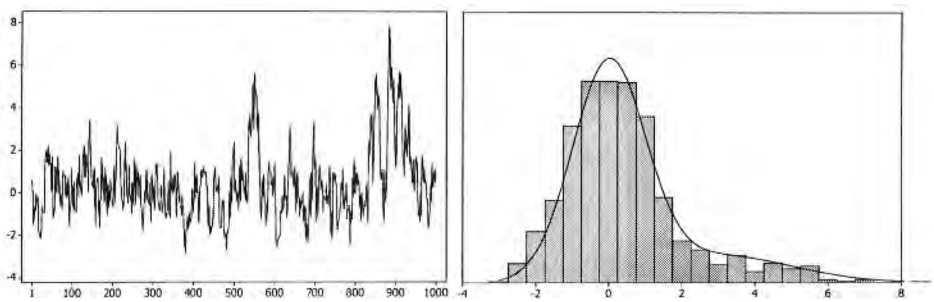

Figure 6.2 Trace plot and histogram of 1000 Metropolis-Hastings values using the random-walk candidate density with standard deviation 1.

for some function $q_2(\theta)$ that must dominate the target in the tails. This is similar to the requirement that the candidate density for acceptance-rejection-sampling must dominate the target in the tails we noted in Chapter 2. We can make sure this requirement is met by graphing logarithms of the target and the candidate density. It is preferable that the candidate density has the same mode as the target as this leads to a higher proportion of accepted candidates. For an independent candidate density, the acceptance probability simplifies to be

$$\begin{aligned}\alpha(\theta,\theta') &= min\left[1, \frac{g(\theta'|y)\,q(\theta',\theta)}{g(\theta|y)\,q(\theta,\theta')}\right]\\ &= min\left[1, \frac{g(\theta'|y)}{g(\theta|y)} \times \frac{q_2(\theta)}{q_2(\theta')}\right].\end{aligned}$$

The acceptance probability is a product of two ratios that give the chain two opposite tendencies. The first ratio $\frac{g(\theta'|y)}{g(\theta|y)}$ gives the chain a tendency to move "uphill" with respect to the target. The second ratio $\frac{q_2(\theta)}{q_2(\theta')}$ gives the chain a tendency to move towards the tails of the candidate density. The acceptance probability depends on the balance of these two opposing tendencies.

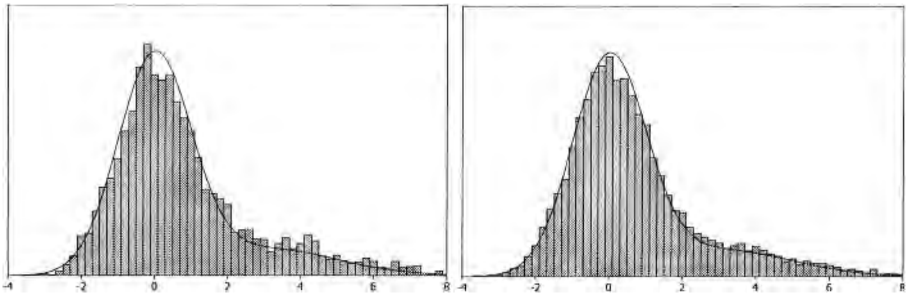

Figure 6.3 Histograms for 5000 and 20000 draws from the Metropolis-Hastings chain using random-walk candidate generating density with standard deviation 1

Table 6.2 Summary of first six draws of the chain using the independent candidate density

Draw	Current value	Candidate	α	u	Accept
1	.4448	-2.7350	.0402	.5871	no
2	.4448	2.3553	.2537	.0987	yes
3	2.3553	.1611	1.000	.7896	yes
5	.1611	-1.5480	.3437	.6357	no
4	.1611	-1.3118	.4630	.1951	yes
6	-1.3118	-2.6299	.1039	.4752	no

Example 8 (continued) *Suppose we use the same unscaled target density given by*

$$g(\theta|y) = .8 \times e^{-\frac{1}{2}\theta^2} + .2 \times \frac{1}{2} e^{-\frac{1}{2\times 2^2}(\theta-3)^2} .$$

This is a mixture of a normal$(0, 1^2)$ *and a normal*$(3, 2^2)$*. Let us use the normal*$(0, 3^2)$ *candidate density as the independent candidate density density. Its shape is given by*

$$q_2(\theta') = e^{-\frac{1}{2\times 3^2}(\theta')^2} .$$

Let the starting value be $\theta = .4448$.

Figure 6.4 shows six consecutive draws from the Metropolis-Hastings chain using the independent candidate distribution along with the unscaled target and the candidate density. Table 6.2 gives a summary of the first 6 draws of this chain. Figure 6.5 shows the traceplot and a histogram for 1000 draws from the Metropolis-Hastings chain using the independent candidate density with mean equal to 0 and standard deviation equal to 3. The independent candidate density allows for large jumps. There may be fewer acceptances than with the random-walk candidate density, however, they will be larger, and we see that the chain is moving through the space very satisfactorily. We see that the histogram has a way to go before it is close to the

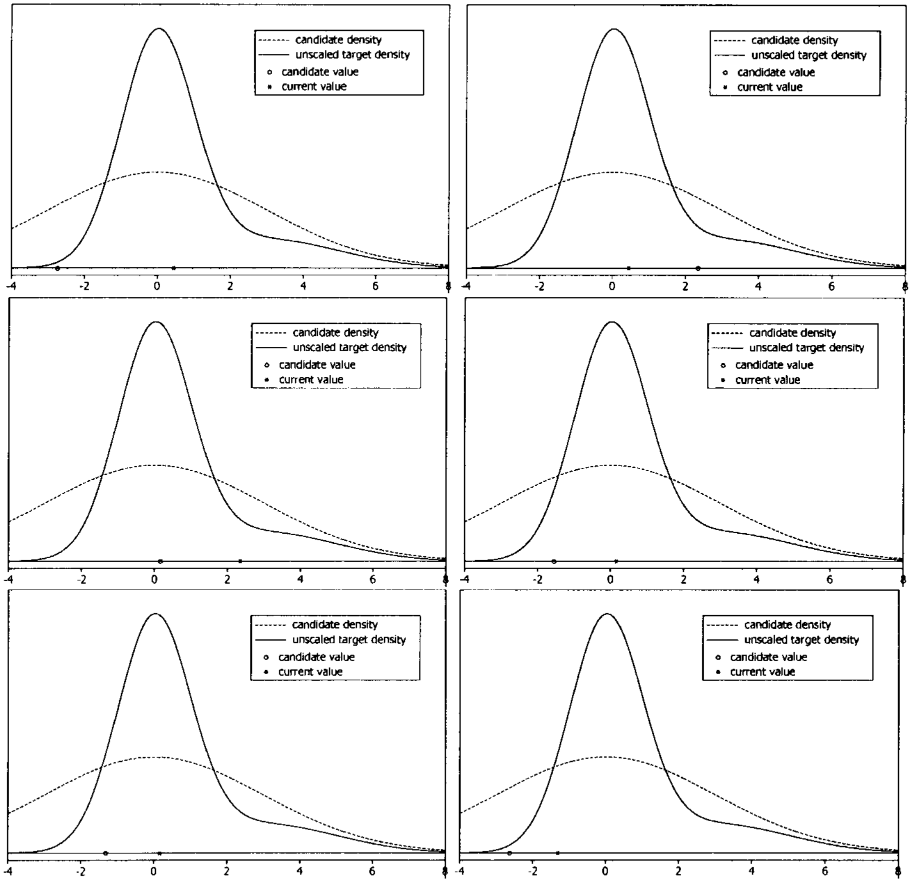

Figure 6.4 Six consecutive draws from a Metropolis-Hastings chain with an independent candidate density. Note: the candidate density remains the same, regardless of the current value.

true posterior. Figure 6.6 shows the histograms for 5000 and 20000 draws from the Metropolis-Hastings chain using the independent candidate density with mean and standard deviation equal to 0 and 3, respectively. We see the chain is getting closer to the true posterior as the number of draws increases.

Note we want to reach the long-run distribution as quickly as possible. There is a trade off between the number of candidates accepted and the distance moved. Generally a chain with a random-walk candidate density will have more candidates accepted, but each move will be a shorter distance. The chain with an independent candidate density will have fewer moves accepted, but the individual moves can be very large. We can see this in comparing the traceplot of the random-walk chain in Figure 6.2 with the trace plot of the independent chain Figure 6.5. The random-walk

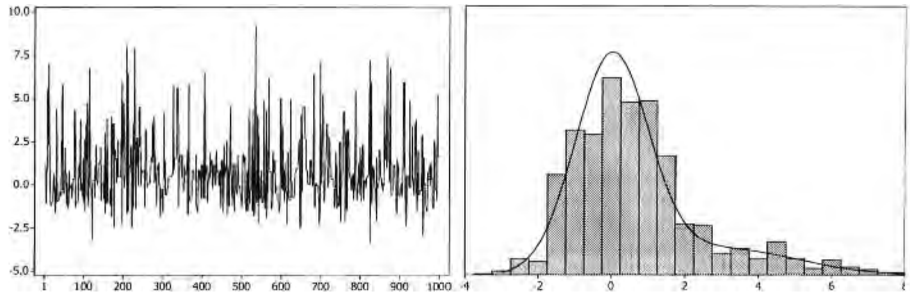

Figure 6.5 Trace plot and histogram of 1000 Metropolis-Hastings values using the independent candidate density with mean 0 and standard deviation 3.

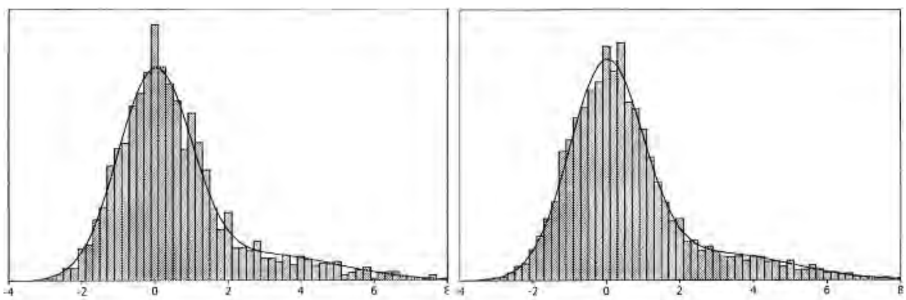

Figure 6.6 Histograms for 5000 and 20000 draws from the Metropolis-Hastings chain using independent candidate generating density with mean 0 and standard deviation 3.

chain is moving more slowly through the whole space, despite more candidates being accepted.

6.2 METROPOLIS-HASTINGS ALGORITHM FOR MULTIPLE PARAMETERS

Suppose we have p parameters $\theta_1, \ldots, \theta_p$. Let the parameter vector be

$$\boldsymbol{\theta} = (\theta_1, \ldots, \theta_p) \,.$$

Let $q(\boldsymbol{\theta}', \boldsymbol{\theta})$ be the candidate density when the chain is at $\boldsymbol{\theta}$ and let $g(\boldsymbol{\theta}|y)$ be the posterior density. The reversibility condition will be

$$g(\boldsymbol{\theta}|y)q(\boldsymbol{\theta}, \boldsymbol{\theta}') = g(\boldsymbol{\theta}'|y)q(\boldsymbol{\theta}', \boldsymbol{\theta})$$

for all states. Most chains won't satisfy the reversibility condition. There will be some values $\boldsymbol{\theta}$ and $\boldsymbol{\theta}'$ where it does not hold. The balance can be restored by