Metropolis Sampling for a 2-Valued Random Variable

Robert J. Mislevy
University of Maryland

November 20, 2001
(Revised December 8, 2001)

Introduction

Suppose we have a random variable X, which can take two values $\{x_1, x_2\}$ with probabilities p and 1-p respectively. That is, $f(x) = p$ if $x=x_1$ and $f(x) = 1-p$ if $x=x_2$. Without loss of generality we presume $p \geq 1-p$. Once stationarity has been achieved, a value from a chain of draws from a Metropolis algorithm (Metropolis & Ulam, 1949) is supposed to have the same distribution as the target density f. What follows is a heuristic argument as to why this is so. The results are then used to offer some intuition behind "burn-in" cycles.

The structure of the argument

We will…

1. Define a symmetric proposal distribution *g,* which gives probabilities of obtaining x_1 or x_2 as a proposal draw given that we are at $x^{(t)} = x_1$ or x_2;
2. Work out acceptance probabilities;
3. Work out the conditional probabilities of $x^{(t+1)}=x_1$ and $x^{(t+1)}=x_2$ given $x^{(t)} = x_1$, and given $x^{(t)} = x_2$;
4. Find the marginal probabilities, or the expected values of obtaining $x^{(t+1)}=x_j$, by averaging the results of (3) with respect to f. It is f again—QED.

The proposal distribution

At Step t of the Metropolis algorithm, we can at either x_1 or x_2. The proposal value y can be either x_1 or x_2. To be symmetric, the proposal distribution must specify, whichever value $x^{(t)}$ is, the same probability π of proposing the current value and 1-π of proposing the other value. That is,

$$\begin{aligned} g\left(y = x_1 \middle| x^{(t)} = x_1\right) &= \pi \\ g\left(y = x_2 \middle| x^{(t)} = x_1\right) &= 1-\pi \\ g\left(y = x_1 \middle| x^{(t)} = x_2\right) &= 1-\pi \\ g\left(y = x_2 \middle| x^{(t)} = x_2\right) &= \pi \end{aligned} \tag{1}$$

Acceptance probabilities

Recall that in Metropolis sampling, the probability of accepting a proposal draw depends on the relative values of the target distribution at the current value $x^{(t)}$ and the proposed value y:

$$\Pr\left(\text{accept } y\right) = \min\left(1, \frac{f\left(y\right)}{f\left(x^{(t)}\right)}\right).$$

Recalling that $f\left(x_1\right) = p \geq \left(1-p\right) = f\left(x_2\right)$, we find that if our proposal is the same place we happen to be already, we stay there; if we are at x_2 and the proposal is x_1, we jump there with probability 1; and if we are at x_1 and the proposal is x_2, we jump there with probability $(1-p)/p$. That is,

$$\begin{aligned}
&\Pr\left(x^{(t+1)} = x_1 \middle| y = x_1, x^{(t)} = x_1\right) = 1 \\
&\Pr\left(x^{(t+1)} = x_2 \middle| y = x_1, x^{(t)} = x_1\right) = 0 \\
&\Pr\left(x^{(t+1)} = x_1 \middle| y = x_2, x^{(t)} = x_1\right) = 1 - f\left(x_2\right)/f\left(x_1\right) = \left(2p-1\right)/p \\
&\Pr\left(x^{(t+1)} = x_2 \middle| y = x_2, x^{(t)} = x_1\right) = f\left(x_2\right)/f\left(x_1\right) = \left(1-p\right)/p \\
&\Pr\left(x^{(t+1)} = x_1 \middle| y = x_1, x^{(t)} = x_2\right) = 1 \\
&\Pr\left(x^{(t+1)} = x_2 \middle| y = x_1, x^{(t)} = x_2\right) = 0 \\
&\Pr\left(x^{(t+1)} = x_1 \middle| y = x_2, x^{(t)} = x_2\right) = 0 \\
&\Pr\left(x^{(t+1)} = x_2 \middle| y = x_2, x^{(t)} = x_2\right) = 1
\end{aligned} \tag{2}$$

Conditional probabilities

What are the conditional probabilities of $x^{(t+1)}=x_1$ and $x^{(t+1)}=x_2$ given $x^{(t)} = x_1$, and given $x^{(t)} = x_2$? They are obtained as the weighted sum of probabilities of getting a value y as a proposal draw, times accepting it as $x^{(t+1)}$, conditional on the value of $x^{(t)}$. Taking each combination in turn, and using values from (1) and (2),

$$\begin{aligned}
&\Pr\left(x^{(t+1)} = x_1 \middle| x^{(t)} = x_1\right) \\
&\qquad = \Pr\left(x^{(t+1)} = x_1 \middle| y = x_1, x^{(t)} = x_1\right)\Pr\left(y = x_1 \middle| x^{(t)} = x_1\right) \\
&\qquad\qquad + \Pr\left(x^{(t+1)} = x_1 \middle| y = x_2, x^{(t)} = x_1\right)\Pr\left(y = x_2 \middle| x^{(t)} = x_1\right) \qquad \text{(3a)} \\
&= 1 \cdot \pi + \left(1 - \frac{1-p}{p}\right)\left(1-\pi\right) \\
&= \pi + \left(\frac{2p-1}{p}\right)\left(1-\pi\right).
\end{aligned}$$

$$\Pr\left(x^{(t+1)}=x_2\,\middle|\,x^{(t)}=x_1\right)$$

$$=\Pr\left(x^{(t+1)}=x_2\,\middle|\,y=x_1,x^{(t)}=x_1\right)\Pr\left(y=x_1\,\middle|\,x^{(t)}=x_1\right)$$

$$+\Pr\left(x^{(t+1)}=x_2\,\middle|\,y=x_2,x^{(t)}=x_1\right)\Pr\left(y=x_2\,\middle|\,x^{(t)}=x_1\right) \quad (3b)$$

$$=\left(\frac{1-p}{p}\right)\pi \;+0\cdot(1-\pi)$$

$$=\left(\frac{1-p}{p}\right)\pi.$$

$$\Pr\left(x^{(t+1)}=x_1\,\middle|\,x^{(t)}=x_2\right)$$

$$=\Pr\left(x^{(t+1)}=x_1\,\middle|\,y=x_1,x^{(t)}=x_2\right)\Pr\left(y=x_1\,\middle|\,x^{(t)}=x_2\right)$$

$$+\Pr\left(x^{(t+1)}=x_1\,\middle|\,y=x_2,x^{(t)}=x_2\right)\Pr\left(y=x_2\,\middle|\,x^{(t)}=x_2\right) \quad (3c)$$

$$=1\cdot(1-\pi)\;+0\cdot\pi$$

$$=(1-\pi).$$

$$\Pr\left(x^{(t+1)}=x_2\,\middle|\,x^{(t)}=x_2\right)$$

$$=\Pr\left(x^{(t+1)}=x_2\,\middle|\,y=x_1,x^{(t)}=x_2\right)\Pr\left(y=x_1\,\middle|\,x^{(t)}=x_2\right)$$

$$+\Pr\left(x^{(t+1)}=x_2\,\middle|\,y=x_2,x^{(t)}=x_2\right)\Pr\left(y=x_2\,\middle|\,x^{(t)}=x_2\right) \quad (3d)$$

$$=0\cdot(1-\pi)\;+1\cdot\pi$$

$$=\pi.$$

Marginal probabilities

What are the marginal probabilities of $x^{(t+1)}=x_1$ and $x^{(t+1)}=x_2$, given $x^{(t)}=x_1$ or $x^{(t)}=x_2$ with probabilities in accordance with f; i.e., p and $(1-p)$ respectively? We find them by averaging the values from (3) accordingly:

$$\begin{aligned}\Pr\left(x^{(t+1)}=x_1\right) &= \Pr\left(x^{(t+1)}=x_1\middle|x^{(t)}=x_1\right)\Pr\left(x^{(t)}=x_1\right)+\Pr\left(x^{(t+1)}=x_1\middle|x^{(t)}=x_2\right)\Pr\left(x^{(t)}=x_2\right)\\ &=\left[\pi+\left(\frac{2p-1}{p}\right)(1-\pi)\right]\cdot p\ +(1-\pi)\cdot(1-p) \qquad (4a)\\ &=p.\end{aligned}$$

$$\begin{aligned}\Pr\left(x^{(t+1)}=x_2\right) &= \Pr\left(x^{(t+1)}=x_2\middle|x^{(t)}=x_1\right)\Pr\left(x^{(t)}=x_1\right)+\Pr\left(x^{(t+1)}=x_2\middle|x^{(t)}=x_2\right)\Pr\left(x^{(t)}=x_2\right)\\ &=\left(\frac{1-p}{p}\right)\pi\cdot p\ +(1-\pi)\cdot(1-p) \qquad (4b)\\ &=1-p.\end{aligned}$$

These are equal to the respective values of f for x_1 and x_2 respectively, which is what we needed to show. One of the "magical" features of the Metropolis algorithm is that the probability π for proposal distribution falls out of the algebra in such a way that the result holds for *any* π between 0 and 1! In our case of a two-valued distribution, it is easy to see that mixing is very slow if π is near 1, since the proposed value is almost always the same as the current value, and no change can occur.

Some intuition for "burn-in"

The preceding section worked out probabilities for a draw from a Metropolis chain assuming stationary had been reached. That is, expected values were calculated for the value of the t+1st value in the series, given that the probabilities for the t^{th} value has the correct probabilities p and 1-p for x_1 and x_2 respectively. Notice that p appears in two ways in the calculations of (4): The first is in the computation of the conditional probabilities of obtaining x_j as the t+1st value given the t^{th} value—these are the terms with the form $\Pr\left(x^{(t+1)}=x_j\middle|x^{(t)}=x_k\right)$. In practice these expressions would be evaluated using the known functional form of f, so the correct values are employed. The second way they are used, though, is in obtaining the expectation of these conditional probabilities, given the probabilities of the possible values of $x^{(t+1)}$ –these are the terms with the form $\Pr\left(x^{(t)}=x_k\right)$.

Suppose instead of the correct value p, which would obtain under stationarity, we were to use a different value p^* for these second terms (i.e., the mixing distribution). (We could think of this, as an example, as the probability distribution from which we chose a starting point $x^{(0)}$ for a Metropolis chain.) Then the probability distribution for the next draw is obtained with the following modification of (4):

$$\Pr^*\left(x^{(t+1)}=x_1\right)$$
$$=\Pr\left(x^{(t+1)}=x_1\middle|x^{(t)}=x_1\right)\Pr^*\left(x^{(t)}=x_1\right)+\Pr\left(x^{(t+1)}=x_1\middle|x^{(t)}=x_2\right)\Pr^*\left(x^{(t)}=x_2\right) \quad (5a)$$
$$=\left[\pi+\left(\frac{2p-1}{p}\right)(1-\pi)\right]\cdot p^* +(1-\pi)\cdot\left(1-p^*\right).$$

$$\Pr^*\left(x^{(t+1)}=x_2\right)$$
$$=\Pr\left(x^{(t+1)}=x_2\middle|x^{(t)}=x_1\right)\Pr^*\left(x^{(t)}=x_1\right)+\Pr\left(x^{(t+1)}=x_2\middle|x^{(t)}=x_2\right)\Pr^*\left(x^{(t)}=x_2\right) \quad (5b)$$
$$=\left(\frac{1-p}{p}\right)\pi\cdot p^* +(1-\pi)\cdot\left(1-p^*\right).$$

Readers can convince themselves that when $p^*<p$,

$$p^* \le \Pr^*\left(x^{(t+1)}=x_1\right)\le p.$$

When $p^*>p$, a similar result holds with the inequalities reversed. In words, when you start a Metropolis chain with the correct calculation of acceptance probabilities but an incorrect specification of the density from which the initial value is drawn, the probability for the next value is closer to the correct probability. Continued draws get you increasingly closer, at a rate that has to do with "mixing." As examples, Table 1 gives values of $\Pr^*\left(x^{(t+1)}=x_1\right)$ calculated with various values of p and p^* and π=.5.

References

Metropolis, N., & Ulam, S. (1949). The Monte Carlo method. *Journal of the American Statistical Association, 44,* 335-341.

Table 1

Values of $\Pr^*\left(x^{(t+1)}=x_1\right)$ for various values of p and p^*, with π=.5.

	p^*								
p	.95	.90	.85	.80	.75	.70	.65	.60	.55
.90	.92	**.90**	.88	.86	.83	.81	.79	.77	.74
.80	.86	.84	.82	**.80**	.78	.76	.74	.73	.71
.70	.77	.76	.74	.73	.71	**.70**	.69	.67	.66
.60	.66	.65	.64	.63	.63	.62	.61	**.60**	.59