Econ 556b | Spring 1999 | C.Sims

Proof of Fixed Point Property for Metropolis Algorithm

We suppose that the target pdf is $p(\theta)$ and that the jump pdf, the pdf from which the new candidate $\tilde{\theta}_1$ is drawn when the previous draw was θ_0, is $f(\tilde{\theta}_1 \mid \theta_0)$. Assume (as is necessary for the Metropolis algorithm to work) that f is symmetric in its two arguments, i.e. that

$$f(x \mid y) = f(y \mid x) \tag{1}$$

for all y and x in the parameter space Θ.

The Metropolis algorithm specifies that the new candidate draw is "accepted" with probability $\pi(\tilde{\theta}_1 \mid \theta_0) = \min\{p(\tilde{\theta}_1)/p(\theta_0), 1\}$, and that otherwise the previous draw θ_0 is repeated. Thus the conditional distribution of $\theta_1 \mid \theta_0$ consists of three components:

i. a discrete weight on the point $\theta_1 = \theta_0$, arising from the possibility of not accepting $\tilde{\theta}_1$ and setting $\theta_1 = \theta_0$ instead;
ii. a component continuously distributed over the set $S(\theta_0) = \{\theta \mid p(\theta) > p(\theta_0)\}$, over which, because in this region $\tilde{\theta}_1$ is always accepted, the pdf is just $f(\theta_1 \mid \theta_0)$;
iii. a component continuously distributed over the set $\Theta - S(\theta_0)$, over which, because the jump may not be accepted, f has to be scaled by the jump probability, yielding as pdf $f(\theta_1 \mid \theta_0) \cdot p(\theta_1)/p(\theta_0)$.

The the probability of the first component is found by cumulating the non-jump probabilities over all of $\Theta - S(\theta_0)$:

$$P(\theta_1 = \theta_0 \mid \theta_0) = \int_{\Theta - S(\theta_0)} \left(1 - \frac{p(\theta)}{p(\theta_0)}\right) f(\theta \mid \theta_0) \, d\theta \, . \tag{2}$$

We wish now to show that if we use this conditional distribution, together with a marginal pdf $h(\theta_0)$ on θ_0 equal to the target pdf p, then the marginal pdf g of θ_1 is also p. The joint distribution of θ_1 and θ_0 when h is the marginal pdf of θ_0 has three components corresponding to the three components of the distribution of $\theta_1 \mid \theta_0$, each formed by multiplying the component of the conditional distribution's pdf by the marginal pdf $h(\theta_0)$. To form g we integrate the two components with densities with respect to θ_0 and add them to the discrete component[1], as in the following equation:

$$g(\theta_1) = \int_{\Theta - S(\theta_1)} h(\theta_1)(1 - \frac{p(\theta)}{p(\theta_1)}) f(\theta \,|\, \theta_1)\, d\theta + \int_{\Theta - S(\theta_1)} f(\theta_1 \,|\, \theta_0) h(\theta_0)\, d\theta_0 + \int_{S(\theta_1)} f(\theta_1 \,|\, \theta_0) \frac{p(\theta_1)}{p(\theta_0)} h(\theta_0)\, d\theta_0 \,. \quad (3)$$

In forming (3), we have used the fact that

$$\{\theta_0 \,|\, \theta_1 \in S(\theta_0)\} = \{\theta_0 \,|\, p(\theta_1) > p(\theta_0)\} = \Theta - S(\theta_1) - \{\theta_1\} \,. \quad (4)$$

Now we use our assumptions of symmetry of f and $h = p$ to simplify (3), arriving at

$$g(\theta_1) = \int_{\Theta - S(\theta_1)} p(\theta_1) f(\theta \,|\, \theta_1)\, d\theta - \int_{\Theta - S(\theta_1)} p(\theta) f(\theta \,|\, \theta_1)\, d\theta + \int_{\Theta - S(\theta_1)} f(\theta_0 \,|\, \theta_1) p(\theta_0)\, d\theta_0 + \int_{S(\theta_1)} f(\theta_0 \,|\, \theta_1) p(\theta_1)\, d\theta_0 \,. \quad (5)$$

Notice that the two middle integrals in (5) are equal and opposite in sign, while the first and last have the same integrand but integrate over two disjoint sets whose union is Θ. But then since $f(\theta_0 \,|\, \theta_1)$ is a pdf as a function of θ_0, its integral over the whole of Θ is one, reducing (5) to $g(\theta_1) = p(\theta_1)$, which is what we have been aiming at.

[1] We are making a "change of variable" in the discrete component, so that in principle we must take account of the Jacobian of the transformation, but here the relation between θ_1 and θ_0 is the identity, so the Jacobian term is just one.