The Validity of Metropolis-Hastings

Sunday, September 09, 2012
1:39 PM

R

- Consider the joint distribution of the draw at time *t-1* and at time *t*:

$$\boxed{f(\theta_{t-1}, \theta_t)} = \cancel{f(\theta_{t-1})g(\theta_t|\theta_{t-1})} \boxed{\frac{f(\theta_t)g(\theta_{t-1}|\theta_t)}{\cancel{f(\theta_{t-1})g(\theta_t|\theta_{t-1})}}} = f(\theta_{t-1})g(\theta_t|\theta_{t-1})$$

$g(\cdot) \to$ proposal density

- Now, set $\theta_{t-1} = x_1$ and $\theta_t = x_2$

$$f(x_1, x_2) = f(x_1)g(x_2|x_1)\frac{f(x_2)g(x_1|x_2)}{f(x_1)g(x_2|x_1)} = \boxed{f(x_2)g(x_1|x_2)}$$

$\theta_t = x_1$
$\theta_{t+1} = x_2$ $\to$ $f(x_2)\, g(x_1|x_2)$

- Obviously: $f(\theta_t, \theta_{t+1})$ takes the same form
- (1) Therefore it must be the case that the density of $f(\theta_t)$ is the same as $f(\theta_{t+1})$ $\to f(\theta_t)\, g(\theta_{t+1}|\theta_t)$ (they're ~~the same given~~ any previous value of the chain)
- (2) Integrate out the proposal density:
  $\int f(\theta_{t-1})g(\theta_t|\theta_{t-1})d\theta_{t-1} = f(\theta_{t-1}) \overset{(1)}{=} f(\theta_t)$
- So… the stationary distribution of the Metropolis-Hastings algorithm is the target density!

A: recovered the target density as the distribution of θ_t

B: the density of $f(\theta_t)$ is the same for all $t(|\theta_{t-1})$

- Still need to check that its resulting density is irreducible and aperiodic
  - Typically true as the number of samples $\to \infty$, but as a practical matter…

  - We have diagnostics for that

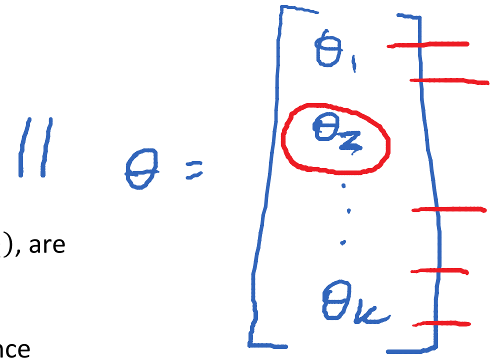
Convergence Diagnostics

Sunday, September 09, 2012
4:35 PM

- Idea: verify that the algorithm is covering the space of possibilities, and is thoroughly mixing
- Simple plot ~~of~~ chain draws: visually assess that chain is moving through the space
- Autocorrelation plot: ensure that the relationship between draws from the chain does not remain high
    - Thinning: take only every *k*th draw from the chain to lower level of autocorrelation
    - Run multiple chains with multiple starting values
- Allow a "burn-in" period

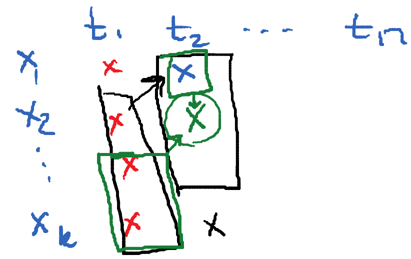
Gibbs Sampler

Sunday, September 09, 2012
5:05 PM



- The Gibbs sampler is a variant of the Metropolis-Hastings algorithm designed for multivariate problems
- Especially useful for cases where the conditional densities, $f(\theta_i|\theta_{-i})$, are analytically known but the joint density is not
- Start k chains for x of length k, then update each element in sequence conditional on all the other elements:

$$x_1^{t+1} \sim f(x_1^{t+1}|x_2^t, x_3^t, \ldots, x_k^t)$$
$$x_2^{t+1} \sim f(x_2^{t+1}|x_1^{t+1}, x_3^t, \ldots, x_k^t)$$
$$\ldots$$
$$x_k^{t+1} \sim f\left(x_k^{t+1}|x_1^{t+1}, x_2^{t+1}, \ldots, x_{k-1}^{t+1}\right)$$
$$x_1^{t+2} \sim f\left(x_1^{t+2}|x_2^{t+1}, x_3^{t+1}, \ldots, x_k^{t+1}\right)$$



- Consider an example: sampling from the multivariate normal distribution
  - As you can find on Wikipedia (among other places), if two variables x_1 and x_2 are distributed multivariate normal with $\mu = \begin{bmatrix} a \\ b \end{bmatrix}$ and covariance matrix $\Sigma = \begin{bmatrix} \sigma_1^2 & \sigma_1\sigma_2\rho \\ \sigma_1\sigma_2\rho & \sigma_2^2 \end{bmatrix}$ the conditional density of $f(x_1|x_2) = \Phi\left(a + \frac{\sigma_1}{\sigma_2}\rho(x_2 - b), (1-\rho^2)\sigma_1^2\right)$ and likewise for x_2.
  - With knowledge of these conditionals, we can use the Gibbs sampler to sample from the multivariate normal density.