Stat 543 2-18-05

Recall

$P[\,i$ iterations required and θ^* $\overbrace{\text{is near } \theta}^{\text{in } "\Delta\theta"}]$

$$\approx \left(1 - \frac{\int L(\theta)g(\theta)\,d\theta}{M}\right)^{i-1} \left(\frac{L(\theta)\,g(\theta)}{M\,h(\theta)}\right) h(\theta)\,\left(\text{vol } "\Delta\theta"\right)$$

$$P[\,\theta^* \text{ is near } \theta\,] = \sum_{i=1}^{\infty} \text{above}$$

$$= \frac{L(\theta)\,g(\theta)\,(\text{vol } "\Delta\theta")}{M} \left(\frac{1}{1 - \left(1 - \frac{\int L(\theta)g(\theta)d\theta}{M}\right)}\right)$$

$$= \frac{L(\theta)\, g(\theta) \,(\text{vol "}\Delta\theta\text{"})}{\int L(\theta)\, g(\theta)\, d\theta}$$

$$\theta^* \sim g(\theta|x)$$

---

to implement the rejection algorithm, I must find appropriate $h(\theta)$ and M — most efficient algorithm is $h(\theta) \approx g(\theta|x)$ and $M \gtrsim 1$ —

$$\frac{L(\theta)\, g(\theta)}{h(\theta)} < M$$

---

another route to using simulation to do Bayes computations is: generate from some other (non-iid) model that produces $\theta_1^*, \theta_2^*, \ldots$

whose empirical dsn is guaranteed to look like $g(\theta|x)$ for large n — one class of models for which this is possible are "Markov chains" — this is MCMC — I'll describe 2 specific versions

Successive Substitution Sampling (Gibbs Sampling)

(implemented in WinBUGS) (see books by Stern/Gelman/Rubin/Carlin or Carlin and Louis or Congdon or ...)

Often the following works (produces a sequence $\theta_1^*, \theta_2^*, \ldots$ whose empirical dsn converges to $g(\theta|x)$):

Start with θ_0^* (possibly generated from approximation to $g(\theta|x)$ or possibly from $g(\theta)$)

With θ_i^* in hand generate θ_{i+1}^* as follows

for each $j = 1, 2, \ldots, k$ in order replace $\theta_{i,j}^*$ by $\theta_{i+1,j}^*$ generate from

$$\underbrace{g\left(\theta_j \,\middle|\, \left(\theta_{i+1,1}^*, \theta_{i+1,2}^*, \ldots, \theta_{i+1,j-1}^*, \theta_{i,j+1}^*, \ldots, \theta_{i,k}^*\right), \text{data}\right)}$$

The posterior dsn of θ_j given all other coordinates of θ evaluated at their current simulated values — each simulation is from a 1-dimensional dsn with density or probability mass function proportional to

$$L(\theta^*_{i+1,1}, \ldots, \theta^*_{i+1,j-1}, \cdot, \theta^*_{i,j+1}, \ldots, \theta^*_{i,k})\, g(\text{same})$$

This could perhaps be done using the rejection algorithm if I can't immediately recognize this form

In practice big issues are

i) making sure your model is not "pathological" and the convergence desired is guaranteed

ii) determining when "transient"/"start-up" effects of simulation have washed out and one should begin accumulating θ^*'s to approximate $g(\theta|x)$

iii) deciding how long to run the simulation

Metropolis – Hastings Algorithm

Another MCMC algorithm

Start with θ_0^* drawn from some dsn that covers the part of Θ where $g(\theta|x)$ puts its mass – for example, in the case where $\Theta \subset \mathbb{R}^k$ a fairly common thing to do is to take

$\hat{\theta}$ an MLE of θ

$$H = -\left(\left. \frac{\partial^2 \log L(\theta)}{\partial \theta_i \partial \theta_j} \right|_{\theta = \hat{\theta}} \right) \qquad H^{-1/2} = \text{matrix square root of } H$$

for $z_1, z_2, \ldots, z_k$ iid $N(0,1)$ (Stern recommends t_4 instead of standard normal)

take

$$Z = \begin{pmatrix} z_1 \\ \vdots \\ z_k \end{pmatrix}$$

use

$$\theta_0^* = H^{-1/2} Z + \hat{\theta}$$

Then, for each $i = 1, 2, \ldots$ let

$$J_i(\theta' \mid \theta)$$

specifies the "jumping dsn" for the ith iteration (the "proposal dsn")

specify for each θ a dsn for θ' over Θ from which I know how to sample – select a candidate

$$\theta_i^{**} \sim J_i(\cdot \mid \theta_{i-1}^{*})$$

" calculate r_i (formula next time) and jump to θ_i^{**} w.p. r_i and stay at θ_{i-1}^{*} w.p. $1-r_i$ "