MCMC I

Markov Chain Monte Carlo Simulation

- Use random numbers to do complex calculations.
- MCMC is NOT medieval mathematics.

High Dimensional Integration

- In Bayesian statistics need to evaluate the integral of the joint probability distribution over all but one of its parameters.
- These integrals can be of very high dimension, say 1000 or 10000 or 100000.
- Numerical integrations requires at least two function evaluations for each dimension – 2^1000 is a big number!

Monte Carlo Simulation

- If we can generate a sample from the joint probability distribution we can use this sample to approximate the Bayesian integrals.
- For example the mean of θ_1

Integral $d\theta_1\ d\theta_{-1}\ \theta_1\ P(\theta_1, \theta_{-1,} Y) \sim \sum \theta_1^{\iota}$

´Equally easy to estimate the mean of θ_1^2 or $\theta_1\theta_2$

- In general can not sample directly from joint probability distribution.

Iterative Simulation

- Generate a sequence of random numbers which converges to a dependent sample from the joint probability distributions.
- Use this dependent sample to approximate the Bayesian integrals.
- Two drawbacks: some computational effort spent reaching convergence, dependent sample contains less information than independent sample.

Markov Chain Simulation

- Construct a markov chain of random numbers. That is the random numbers generated only depend on the current state of the chain.
- In certain cases this chain of random numbers will converge to a dependent sample from the joint probability distribution.

The Gibbs Sampler

- For each stochastic variable in the Bayesian model replace its value with a random number sampled from the full conditional distribution.
- Repeat procedure many times.
- After some iterations use the sample of random numbers for inference.

Gibbs Sampler continued

- Easy algorithm to think about.
- Exploits the factorisation properties of the joint probability distribution.
- No difficult choices to be made to tune algorithm.
- Can be difficult (impossible) to sample from the full conditional distributions.

Metropolis Hastings Sampler

- For each stochastic variable θ in the Bayesian model associate a proposal distribution $\pi(x, \theta)$. Draw a sample $\theta 1$ from π and replace the value of θ with $\theta 1$ if

$$\alpha = \min(P(\theta 1)\pi(\theta, \theta 1) \;/\; P(\theta)\pi(\theta 1, \theta), 1)$$

is greater than $U(0,1)$ where $P(\theta)$ is the full conditional for θ.

α is called the acceptance ratio.

The proposal distribution $\pi(x, \theta)$

- The π is (can be) different for each variable.
- The π need not depend on θ. Can have $\pi(x, \theta) = \pi(x)$ this is called an independence sampler.
- If we choose $\pi(x, \theta)$ to be $P(x)$ this independence sampler is the Gibbs Sampler.
- If π is chosen such that $\pi(x, \theta) = \pi(\theta, x)$ then acceptance ratio simplifies (π cancels out).

$\pi(x, \theta)$ continued

- A common choice for $\pi(x, \theta)$ is $\pi(|x - \theta|)$ this is called a random walk sampler.

Problems with Metropolis Hastings Sampling

- How should the proposal distribution be chosen?
- If the proposal distribution depends on some parameters how should these parameters be chosen? For example if

  $\pi(x, \theta) = \mathrm{Normal}(x|\theta, \tau)$ what value should τ have.

Slice Sampling

- A new sampling algorithm due to Radford Neal.
- Evaluate the full conditional distribution at the current point θ and multiply it by a standard uniform (ie P(θ)U(0,1)). Calculate the two values x_1, x_2 such that

$$P(x_i) = P(\theta)U(0,1) \quad I = 1, 2$$

Replace θ by a sample from U(x1, x2).