Stochastic Simulation Markov Chain Monte Carlo

Bo Friis Nielsen

Institute of Mathematical Modelling Technical University of Denmark 2800 Kgs. Lyngby – Denmark Email: bfni@dtu.dk

MCMC: What we aim to achieve



DIU

We have a variable X with a "complicated" distribution.

We cannot sample X directly.

We aim to generate a sequence of X_i 's

- which each has the same distribution as \boldsymbol{X}
- but we (have to) allow them to be dependent.

This is an **inverse problem** relative to what we just discussed and to the queueing exercise:

We start with the distribution of X, and aim to design a state machine which has this steady-state distribution.

MCMC example from Bayesian statistics Prior distribution of parameter

 $P \sim U(0,1)$: $f_P(p) = \mathbf{1} \quad (0 \le p \le 1)$

Distribution of data, conditional on parameter

X for given P = p is Binomial(n, p)

i.e. the data has the conditional probabilities

$$\mathsf{P}(X=i|P) = \left(\begin{array}{c}n\\i\end{array}\right) P^{i}(1-P)^{n-i}$$

DIU

The posterior distribution of P



Conditional density of parameter, given observed data X = i (the posterior distribution):

$$f_{P|X=i}(p) = f_P(p) \frac{\mathsf{P}(X=i|P=p)}{\mathsf{P}(X=i)} c f_P(p) \mathsf{P}(X=i|P=p)$$

We need the unconditional probability of the observation:

$$\mathsf{P}(X=i) = \int_0^1 f_P(p) \left(\begin{array}{c}n\\i\end{array}\right) p^i (1-p)^{n-i} dp$$

We *can* evaluate this; but in more complex models we might not. AIM: To sample from $f_{P|X=i}$, without evaluating c = 1/P(X = i).

The posterior distribution





When to apply MCMC?

The distribution is given by

$$f(x) = c \cdot g(x)$$

DTU

where the *unnormalized density* g can be evaluated, *but* the normalising constant c cannot be evaluated (easily).

$$c = \frac{1}{\int_{\mathbf{X}} g(x) \, dx}$$

This is frequently the case in Bayesian statistics - the posterior density is proportional to the likelihood function

Note (again) the similarity between simulation and evaluation of integrals

DIU

02443 - lecture 8

When to apply MCMC? - continued

We want to sample from a distribution is given by

$$f(x) = c \cdot g(x)$$

DTU

DIU

where the *unnormalized density* g can be evaluated, *but* the normalising constant c cannot be evaluated (easily).

$$c = \frac{1}{\int_{\mathbf{X}} g(x) \, dx}$$

We generate samples from a Markov chain, where we can prove, that the limiting (invariant) distribution is our target distribution (f(x))

Metropolis-Hastings algorithm



- Proposal distribution $h({m x},{m y})$
- Acceptance of solution? The solution will be accepted with probability

$$\min\left(1, \frac{f(\boldsymbol{y})h(\boldsymbol{y}, \boldsymbol{x})}{f(\boldsymbol{x})h(\boldsymbol{x}, \boldsymbol{y})}\right) = \min\left(1, \frac{g(\boldsymbol{y})h(\boldsymbol{y}, \boldsymbol{x})}{g(\boldsymbol{x})h(\boldsymbol{x}, \boldsymbol{y})}\right)$$

- Avoiding the troublesome constant *c*!
- Frequently we apply a symmetric proposal distribution $h(\boldsymbol{y}, \boldsymbol{x}) = h(\boldsymbol{x}, \boldsymbol{y})$ Metropolis algorithm to get $\left(= \min\left(1, \frac{g(\boldsymbol{y})}{g(\boldsymbol{x})}\right)$ for $h(\boldsymbol{y}, \boldsymbol{x}) = h(\boldsymbol{x}, \boldsymbol{y}) \right)$

D I U

Metropolis Hastigs algorithm and local balance The transition rate q(x, y) from x to y and vice versa is

$$q(\boldsymbol{x}, \boldsymbol{y}) = h(\boldsymbol{x}, \boldsymbol{y}) \min\left(1, \frac{g(\boldsymbol{y})h(\boldsymbol{y}, \boldsymbol{x})}{g(\boldsymbol{x})h(\boldsymbol{x}, \boldsymbol{y})}\right)$$

and

$$q(\boldsymbol{y}, \boldsymbol{x}) = h(\boldsymbol{y}, \boldsymbol{x}) \min\left(1, \frac{g(\boldsymbol{x})h(\boldsymbol{x}, \boldsymbol{y})}{g(\boldsymbol{y})h(\boldsymbol{y}, \boldsymbol{x})}\right)$$

Suppose $g(\boldsymbol{y})h(\boldsymbol{y},\boldsymbol{x}) < g(\boldsymbol{x})h(\boldsymbol{x},\boldsymbol{y})$ then

$$f(\boldsymbol{x})q(\boldsymbol{x},\boldsymbol{y}) = cg(\boldsymbol{x})h(\boldsymbol{x},\boldsymbol{y})\frac{g(\boldsymbol{y})h(\boldsymbol{y},\boldsymbol{x})}{g(\boldsymbol{x})h(\boldsymbol{x},\boldsymbol{y})} = cg(\boldsymbol{y})h(\boldsymbol{y},\boldsymbol{x}) = f(\boldsymbol{y})q(\boldsymbol{y},\boldsymbol{x})$$



g

Random Walk Metropolis-Hastings

A simple symmetric proposal distribution is the random walk DTU

- 1. At iteration i, the state is X_i
- 2. Propose to jump from X_i to $Y_i = X_i + \Delta X_i$ where ΔX_i is sampled independently from a symmetric distribution
 - If $g(Y) \ge g(X_i)$, accept
 - If $g(Y) \leq g(X_i)$, accept w.p. $g(Y)/g(X_i)$
- 3. On accept: Set $X_{i+1} = Y_i$ and goto 1.
- 4. On reject: Set $X_{i+1} = X_i$ and goto 1.

DIU

 Ξ

Proposal distribution (Gelman 1998)



DH

11

- A good proposal distribution has the following properties
 - \diamond For any $oldsymbol{x}$, it is easy to sample from $h(oldsymbol{x},oldsymbol{y})$
 - It is easy to compute the accpetance probability
 - Each jump goes a reasonable distance in the parameter space
 - The proposals are not rejected too frequently



A new proposal can be anywhere in the full region



A new proposal can be anywhere in the full region. However, typically it will be in the vicinity of the current

We generate proposal and do acceptance/rejection for one for one dimension at a time. It can be done systematically or random.

DTU



Possibilities in x-direction



DTU

₩

Accepted candidate in x-direction





Possibilities in *y*-direction



DTU

₩

Accepted candidate in *y*-direction





Final update

Gibss sampling

DTU

19

- Applies in multivariate cases where the conditional distribution among the coordinates are known.
- For a multidimensional distribution x the Gibss sampler will modify only one coordinate at a time.
- Typically d-steps in each iteration, where d is the dimension of the parameter space , that is of $m{x}$



DTU

 Ξ

 Each dimension is updated at a time. According to the <u>conditional distribution. Here the first dimension</u> <u>02443 - lecture 8</u>



Draw from conditional distribution (no acceptance test).



Possibilities in y-direction





Final state after updates in both directions.

Different perspective modelling with Markov chains as oppposed to MCMC sampling

- For an ordinary Markov chain we know $I\!\!P$ and find π analytically or by simulation
- When we apply MCMC
 - For a discrete distribution we have π = cα construct P which has no physical interpretation in general and obtain samples from π by simulation
 - ♦ For a continuous distribution we have the density f(x) = cg(x), construct a transition kernel P(x, y) and get samples from f(x) by simulation.

 $D \mid U$

25

Remarks

- The method is computer intensitive
- It is hard to verify the assumptions (Read: impossible)
- Warmup period strongly recommended (necessary indeed!)
- The samples are dependent (typically correlated)
- Should be run several times with different starting conditions
 - Comparing within run variance with between run variance
- Check the BUGS site:

http://www.mrc-bsu.cam.ac.uk/bugs/and/or links given at the BUGS site



DIU

Further reading



 \mathbf{D}

27

- A. Gelman, J.B. Carlin, H.S. Stern, D.B. Rubin: Bayesian Data Analysis, Chapmann & Hall 1998, ISBN 0 412 03991 5
- W.R. Gilks, S. Richarson, D.J. Spiegelhalter: Markov chain Mone Carlo in practice, Chapmann & Hall 1996, ISBN 0 412 05551 1

Beyond Random Walk Metropolis-Hastings

- Proposed points Y_i can be generated with other schemes this would change the acceptance probabilities.
- In mulitvariate situations, we can process one co-ordinate at a time
- If we know conditional distributions in the mulitvariate setting, then we can apply Gibbs sampling
- This is well suited for *graphical models* with many variables, which each interact only with a few others
- (Decision support systems is a big area of application)
- Many hybrids and specialized versions exist
- Very active research area, both theory and applications

Exercise 6: Markov Chain Monte Carlo

1. The number of busy lines in a trunk group (Erlang system) is given by a truncated Poisson distribution

$$P(i) = c \cdot \frac{A^i}{i!}, \quad i = 0, \dots m$$

DTU

Generate values from this distribution by applying the Metropolis-Hastings algorithm, verify with a χ^2 -test. You can use the parameter values from exercise 4.

2. For two different call types the joint number of occupied lines is given by

$$P(i,j) = c \cdot \frac{A_1^i}{i!} \frac{A_2^j}{j!} \qquad 0 \le i+j \le m$$

You can use $A_1, A_2 = 4$ and m = 10. (a) Use Metropolis-Hastings, directly to generate variates from this distribution.

- (b) Use Metropolis-Hastings, coordinate wise to generate variates from this distribution.
- (c) Use Gibbs sampling to sample from the distribution. This is (also) coordinate-wise but here we use the exact conditional distributions. You will need to find the conditional distributions analytically.

In all three cases test the distribution fit with a χ^2 test

The system can be extended to an arbitrary dimension, and we can add restrictions on the different call types.

3. We consider a Bayesian statistical problem. The observations are $X_i \sim N(\Theta, \Psi)$, where the prior distribution of the pair $(\Xi, \Gamma) = (\log(\Theta), \log(\Psi))$ is standard normal with correlation $\rho = \frac{1}{2}$. The joint density f(x, y) of (Θ, Ψ) is

$$f(x,y) = \frac{1}{2\pi x y \sqrt{1-\rho^2}} e^{-\frac{\log(x)^2 - 2\rho \log(x) \log(y) + \log(y)^2}{2(1-\rho^2)}}$$

which can be derived using a standard change of variable technique. The task of this exercise is now to sample from the posterior distribution of (Θ, Ψ) using Markov Chain Monte Carlo.

- (a) Generate a pair (θ, ψ) from the prior distribution, i.e. the distribution for the pair (Θ, Ψ) , by first generating a sample (ξ, γ) of (Ξ, Γ) .
- (b) Generate $X_i = 1, ..., n$ with the values of (θ, ψ) you obtained in item 3a. Use n = 10.
- (c) Derive the posterior distribution of (Θ, Ψ) given the sample. **Hint 1** Apply Bayes theorem in the density version.

Hint 2 The sample mean and sample variance are independent. The sample mean follows a normal distribution, while a scaled version of the sample variance follows a χ^2 distribution. This can be used to simplify the expression.

- (d) Generate MCMC samples from the posterior distribution of (Θ, Ψ) using the Metropolis Hastings method.
- (e) Repeat item 3d with n = 100 and n = 1000, still using the values of (θ, ψ) from item 3a. Discuss the results.