Introduction to MCMC
Lucy M. Li, PhD
Introduction
Theory
Interpretation
Working example
Practical considerations
Conclusions

Introduction to MCMC

Lucy M. Li, PhD

May 9 & 11, 2018

1. Introduction
2. Theory
3. Interpretation
4. Working example
5. Practical considerations
6. Conclusions

Introduction

Parameter estimation

- Important epidemiological parameters help to inform public health policies aiming to control, prevent, and eliminate infectious diseases

How can we estimate parameters?

1. Identify relevant **parameters** e.g. R_0, vaccine efficacy, duration of infection
2. Formulate mathematical **model** that describes disease transmission and characterizes the relationship between parameters and observations
3. Estimate parameter values by **fitting** models to data

Challenges of model fitting

- Incomplete data
- Temporally correlated data
- Multiple parameters need to be estimated at the same time
- Often no analytical solution to likelihood functions or posterior distributions

How can Markov chain Monte Carlo help?

- An inference framework that is particularly useful for Bayesian analyses by approximating the posterior distribution through sampling
- Efficient algorithm to explore parameter space
- Enables imputation of missing data and parameter estimation at the same time
- Flexible framework that works with a variety of data including time-series data and phylogenetic data

Examples of MCMC in literature

- Hopefully you'll understand these sentences by the end of the class!

Fraser et al (2009) *Science*.: an early analysis of pandemic H1N1

'In a Bayesian framework, the joint posterior distribution of the "real" number of cases M_t for $t = 1$ to T and of the individual reproduction number R_t for $t = 1$ to T is explored via Markov Chain Monte Carlo sampling'

Morton and Finkenstadt (2005) *Applied Statistics*.: analysis of measles

'To estimate all parameters and states simultaneously of the model. . . we use MCMC sampling. This involves simulation from the joint posterior density by setting up a Markov chain whose stationary distribution is equal to this target posterior density.'

Bhatt et al () *Nature*.: estimating global burden of dengue

Theory

An analogy for MCMC

An

Markov chain

Components

1. A series of random variables $X_{t=0}, ..., X_{t=T}$ with values in state space $S = \{s_1, ..., s_n\}$.
2. Transition probabilities are described by $\mathbf{M} = \{m_{ij}\}$ where $m_{ij} = Pr(X_{t+1} = s_j | X_t = s_i)$, and $m_{ij} \geq 0$ for all i, j.
3. Initial probability distribution of state values is defined by $\nu^{(0)}$, where $\nu_i^{(0)} = Pr(X_0 = s_i)$.

Key Markovian property

Probabilities state values at time $t+1$ only depends on the state value at the previous time step t:
$Pr(X_{t+1} = s_j | X_t) = Pr(X_{t+1} = s_j | X_t, X_{t-1}, ..., X_0)$

Stationary distribution

Some Markov chains have a stationary distribution $\nu^{(n)}$ such that $\nu^{(n)} = \nu^{(n)} M^n$ as $n \to \infty$.

Markov chain: modeling the weather

Components

1. Weather on consecutive days $X_{t=0}, \dots, X_{t=T}$ with values in state space $S = \{s_1 = \text{Sunny}, s_2 = \text{Cloudy}, s_3 = \text{Rainy}\}$.
2. Transition probabilities are described by $\mathbf{M} = \{m_{ij}\}$

$$\mathbf{M} = \begin{array}{c} \\ \text{Sunny} \\ \text{Cloudy} \\ \text{Rainy} \end{array} \begin{array}{c} \begin{array}{ccc} \text{Sunny} & \text{Cloudy} & \text{Rainy} \end{array} \\ \begin{pmatrix} 0.55 & 0.35 & 0.10 \\ 0.30 & 0.40 & 0.30 \\ 0.20 & 0.40 & 0.40 \end{pmatrix} \end{array}$$

3. Initial probability distribution of weather is defined by

$$\nu^{(0)} = \begin{array}{c} \text{Sunny} \\ \text{Cloudy} \\ \text{Rainy} \end{array} \begin{pmatrix} \frac{1}{3} \\ \frac{1}{3} \\ \frac{1}{3} \end{pmatrix}.$$

Stationary distribution of Markov Chains π

What is M^t as $t \to \infty$? $M^2 = M \cdot M$

$$\mathbf{M}^2 = \begin{array}{c c} & \begin{array}{ccc} \text{Sunny} & \text{Cloudy} & \text{Rainy} \end{array} \\ \begin{array}{c} \text{Sunny} \\ \text{Cloudy} \\ \text{Rainy} \end{array} & \begin{pmatrix} 0.43 & 0.37 & 0.20 \\ 0.35 & 0.39 & 0.27 \\ 0.31 & 0.39 & 0.30 \end{pmatrix} \end{array}$$

$$\begin{aligned} M_{ij}^2 &= Pr(X_{t+1} = j | X_{t-1} = i) \\ &= \sum_{k=1}^{n} Pr(X_{t+1} = j | X_t = k) \cdot Pr(X_t = k | X_{t-1} = i) \end{aligned}$$

Stationary distribution of Markov Chains π

$$\mathbf{M}^3 = \begin{array}{c c} & \begin{array}{ccc} \text{Sunny} & \text{Cloudy} & \text{Rainy} \end{array} \\ \begin{array}{c} \text{Sunny} \\ \text{Cloudy} \\ \text{Rainy} \end{array} & \begin{pmatrix} 0.39 & 0.38 & 0.23 \\ 0.36 & 0.38 & 0.26 \\ 0.35 & 0.38 & 0.27 \end{pmatrix} \end{array}$$

$$\mathbf{M}^4 = \begin{array}{c c} & \begin{array}{ccc} \text{Sunny} & \text{Cloudy} & \text{Rainy} \end{array} \\ \begin{array}{c} \text{Sunny} \\ \text{Cloudy} \\ \text{Rainy} \end{array} & \begin{pmatrix} 0.37 & 0.38 & 0.25 \\ 0.36 & 0.38 & 0.25 \\ 0.36 & 0.38 & 0.26 \end{pmatrix} \end{array}$$

Markov chain

Markov chain Monte Carlo

- In MCMC algorithms, we take advantage of this property by using the stationary distribution $\nu^{(n)}$ to estimate probability distributions of interest such as the posterior distribution $P(\theta|D)$)

Overview of MCMC

1. Initialize parameter values θ_{init}.
2. Propose a new parameter values θ_{new} given θ_{init}.
3. Accept or reject the new values probabilistically based on prior and likelihood.
4. If θ_{new} is accepted, set $\theta_{\text{init}} := \theta_{\text{new}}$.
5. Repeat steps 2-4 until sufficient samples of θ have been obtained.

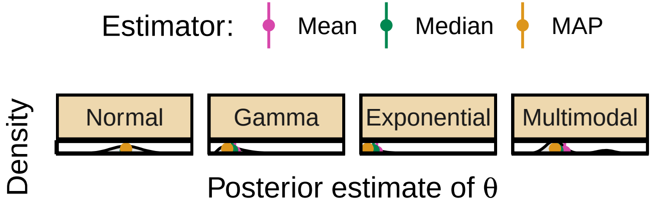
Animation

https://chi-feng.github.io/mcmc-demo/app.html#RandomWalkMH,standard

Interpretation

Summarizing parameter estimates

- Maximum a posteriori probability (MAP) estimate is the **mode** of the samples from the posterior distribution
- More common to report the **median** or **mean**



Quantifying uncertainty

- Standard deviation - doesn't account for skew
- Quantile interval - captures uncertainty of over-dispersed distribution better than the SD, but may over-estimate the size of the credible interval
- 95% highest posterior density (HPD) region useful for non-Gaussian posterior distribution

Working example

Code

Use Swirl

Practical considerations

Parameter transformations

- positivity (log)
- bounds (logit)

Burn-in period

- At the beginning, the chain samples from the prior

Convergence diagnostics

- Start multiple chains and use Gelman-Rubins diagnostic
- Check autocorrelation plot
- Cannot be completely sure that the chain has not converged to a local minimum
- VISUAL CHECK!!!!

Proposal distribution

- Adaptive MCMC
- Joint proposal distribution

Initialization

- Multiple chains starting at different points
- Latin-hypercube sampling

Multi-modal distribution

- Often occurs when parameters are correlated

Computational

- Thinning to store a subset of sampled values, because chain is auto-correlated
- Debugging: run with just the prior distribution to check that the chain is sampling from the posterior

Conclusions

Further reading - More theoretical papers/books on MCMC in epidemiology

- Metropolis et al (1953) Equation of State Calculations by Fast Computing Machines. *J. Chem. Phys.*
- Hastings (1970) Monte Carlo Sampling Methods Using Markov Chains and Their Applications. *Biometrika*.
- Gilks et al (1995) MCMC in practice - Brooks et al (2011) Handbook of MCMC
- Gibson and Renshaw (1998)
- O'Neill and Roberts (1999)
- O'Neill (2002) A tutorial introduction to Bayesian inference for stochastic epidemic models using Markov chain Monte Carlo methods - Statistical inference in a stochastic epidemic SEIR model with control intervention: Ebola as a case study
- Morton and Finkenstädt (2005) Discrete time modelling of disease incidence time series by using Markov chain Monte Carlo methods. *Applied Statistics*.

Further reading - Applications of MCMC to infectious disease epi

- Cauchemez et al (2004) A Bayesian MCMC approach to study transmission of influenza: application to household longitudinal data
- Fraser et al (2009) Pandemic potential of a strain of influenza A (H1N1): early findings. *Science*. [Most relevant section is in Supporting Online Material Section 2]
- van Boven et al (2010) Estimation of measles vaccine efficacy and critical vaccination coverage in a highly vaccinated population. *J. R. Soc. Interface*.

Further reading - MCMC in genetics

Newer developments