Bayesian Analysis via MCMC

Jingqi Zhu

The University of Manchester

June 7, 2022

Overview

Bayesian analysis and its primary challenge

- Bayesian analysis: a statistical inference paradigm combining prior and data based on Bayes' theorem.
- Objective: posterior distribution $p(\boldsymbol{\theta}|\boldsymbol{x})$

$$p(\boldsymbol{\theta}|\boldsymbol{x}) = \frac{p(\boldsymbol{x}|\boldsymbol{\theta})p(\boldsymbol{\theta})}{p(\boldsymbol{x})} \propto p(\boldsymbol{x}|\boldsymbol{\theta})p(\boldsymbol{\theta}) \tag{1}$$

where $p(\boldsymbol{x}) = \int_{\boldsymbol{\Theta}} p(\boldsymbol{x}|\boldsymbol{\theta})p(\boldsymbol{\theta})d\boldsymbol{\theta}$ is a normalizing constant.

- Moreover, most summary statistics can be expressed as posterior expectations of $f(\theta)$

$$E[f(\boldsymbol{\theta})|\boldsymbol{x}] = \frac{\int_{\boldsymbol{\Theta}} f(\boldsymbol{\theta})p(\boldsymbol{x}|\boldsymbol{\theta})p(\boldsymbol{\theta})d\boldsymbol{\theta}}{\int_{\boldsymbol{\Theta}} p(\boldsymbol{x}|\boldsymbol{\theta})p(\boldsymbol{\theta})d\boldsymbol{\theta}} \tag{2}$$

- Challenge: approximating integrals/expectations.

Bayesian analysis and its primary challenge

- Task: Approximate $E_\pi[f(\boldsymbol{\theta})] = \int f(\boldsymbol{\theta})\pi(\boldsymbol{\theta})d\boldsymbol{\theta}$
- Numerical integration (trapezium/Simpson's method, etc) suffers the curse of dimensionality: $e \sim O(n^{-\frac{1}{d}})$.
- Monte Carlo integration: sample iid $\boldsymbol{\theta}_i \sim \pi$, $1 \leq i \leq n$

$$\hat{I}(n) = \frac{1}{n}\sum_{i=1}^{n} f(\boldsymbol{\theta}_i) \tag{3}$$

  Converge faster: $e \sim O(n^{-\frac{1}{2}})$.
- New challenge: sampling from a target distribution $\pi(\boldsymbol{\theta})$.

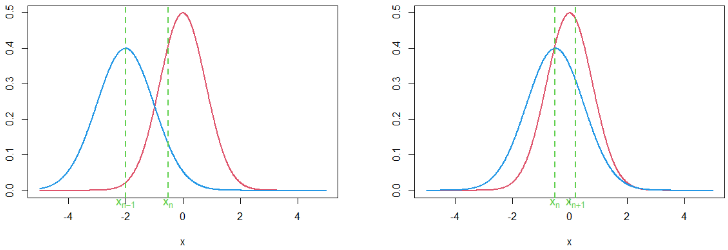
Classical simulation methods

- Classical simulation methods fail to sample an arbitrary target posterior π in high dimension.

Methods	Problem
Inverse cdf method	no available cdf for posterior
Transformation method	only work for standard distributions
Accept-Reject method	hard to find high dimensional proposal
Importance sampler	hard to find high dimensional proposal

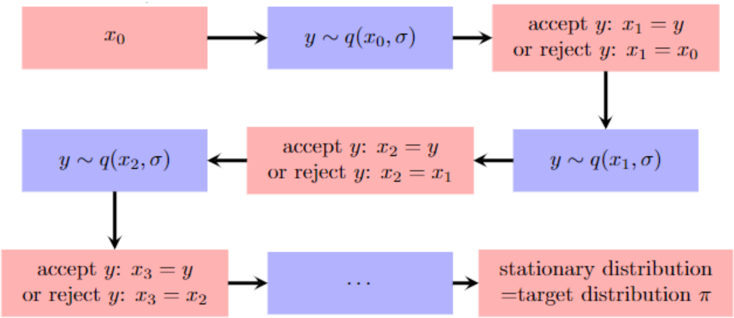
Rethinking 1 - dependence

- Rethinking 1: Monte Carlo requires independent samples. Side effect: erase information from previous sample when drawing new samples!
- Independency-Efficiency tradeoff: improve the efficiency of sampling by allowing some dependence among samples?
- Rather than universal proposal, use moving proposal!



Rethinking 2 - convergence

- Rethinking 2: Hard to sample target distribution directly. Generate a sequence converging to target distribution?
- Markov chain converging to stationary distribution which happens to be the target distribution.
- Markov chain+Monte Carlo=Markov chain Monte Carlo.



Metropolis-Hastings algorithm

- 1. starting from an initial point x_0
  2. draw candidate from proposal $Y \sim q(Y|X_i)$
  3. accept candidate with probability

$$\alpha(Y|X_i) = min\{\frac{\pi(Y)q(X_i|Y)}{\pi(X_i)q(Y|X_i)}, 1\} \tag{4}$$

- Acceptance rate $\bar{\alpha} = \lim_{N\to\infty} \frac{1}{N} \sum_{i=0}^{N} \alpha(Y|X_i)$
- Only requires knowing target up to a constant.

Metropolis-Hastings algorithm

- Proposal should be close to the target as possible
- We usually choose proposals with larger variance or heavier tails than the target to ensure the whole space is explored.
- Parameters of proposal can be tuned according to acceptance rate, trace plot and autocorrelation plot.
- Independent Metropolis-Hastings

$$\alpha(Y|X_i) = min\{\frac{\pi(Y)q(X_i)}{\pi(X_i)q(Y)}, 1\} \tag{5}$$

- Random Walk Metropolis

$$\alpha(Y|X_i) = min\{\frac{\pi(Y)}{\pi(X_i)}, 1\} \tag{6}$$

Ideal random walk proposal should have acceptance rate approximately 25%.

Gibbs sampling

- 1. determine an initial point $\boldsymbol{x}^{(0)}$

  2. generate $X_1^{(i+1)} \sim \pi(x_1 | x_2^{(i)}, \ldots, x_d^{(i)})$

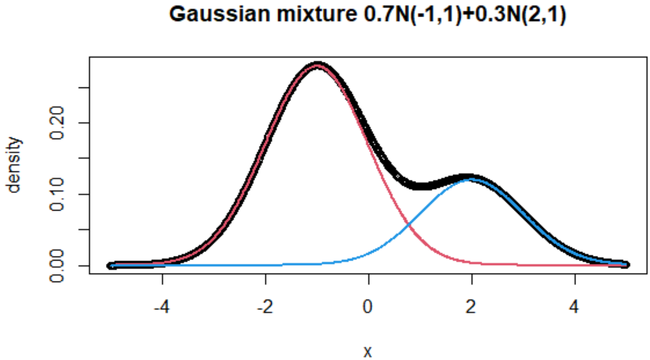
     generate $X_2^{(i+1)} \sim \pi(x_2 | x_1^{(i+1)}, x_3^{(i)}, \ldots, x_d^{(i)})$

     ...

     generate $X_d^{(i+1)} \sim \pi(x_d | x_1^{(i+1)}, x_2^{(i+1)}, \ldots, x_{d-1}^{(i+1)})$
- Even though joint distribution can be complicated, full conditionals are univariate, thus Gibbs sampling is very efficient.
- Gibbs sampling accept every move, no need to concern acceptance rate or tuning parameters.

Algorithm comparison

- Estimate parameters of Gaussian mixture model $pN(\mu_1, 1/\tau_1) + (1-p)N(\mu_2, 1/\tau_2)$.
- Underlying true model is $0.7N(-1,1) + 0.3N(2,1)$.
- Non-informative priors: $\mu_1 \sim N(-1,4), \mu_2 \sim N(2,4)$, $\tau_1 \sim Exp(1), \tau_2 \sim Exp(1)$, $p \sim Beta(3,2)$.



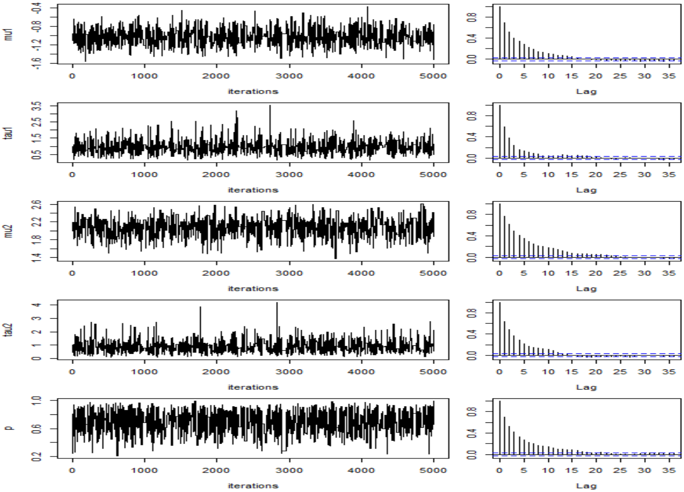
Algorithm comparison

- Independent proposals: $\mu_1 \sim N(-1, \frac{var(x)}{n}), \mu_2 \sim N(2, \frac{var(x)}{n}),$ $\tau_1 \sim Gamma(2,2), \tau_2 \sim Gamma(2,2),\ p \sim Beta(7,3).$
- Random proposals: for $\mu_1, \mu_2, \tau_1, \tau_2$ are $N(0, 0.25^2)$, for p is $N(0, 0.005^2)$.
- Gibbs: latent variables, data argumentation.

	mu1	tau1	mu2	tau2	p
IndepMH	-0.91	1.10	1.96	1.53	0.66
RWM	-1.02	0.96	1.32	0.73	0.61
Gibbs	-1.06	0.99	2.84	0.92	0.70
True Model	-1	1	2	1	0.70

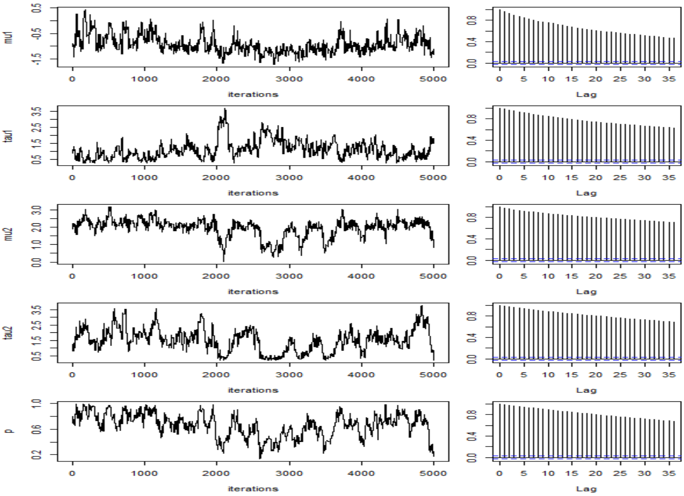
Independent Metropolis-Hastings

- Acceptance rate: 0.26.

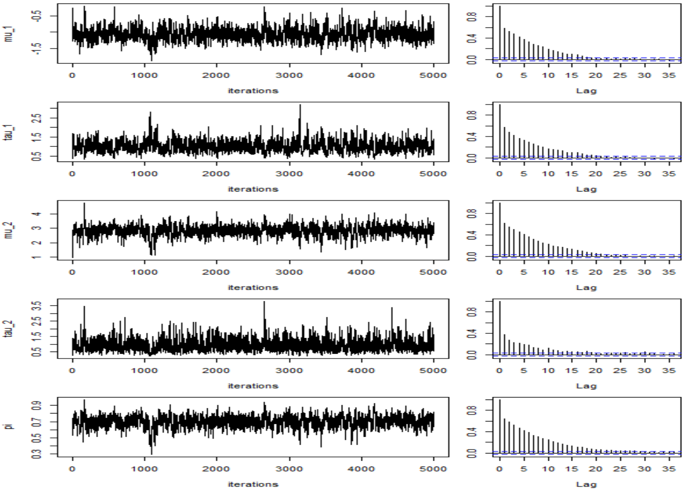


Random walk Metropolis

- Acceptance rate: 0.25.



Gibbs sampling



Summary and bigger picture