Simulation laboratory 5: Markov chain Monte Carlo methods

Pavel Ilinov

Transport and Mobility Laboratory
School of Architecture, Civil and Environmental Engineering
École Polytechnique Fédérale de Lausanne

28 March 2024

Overview

Objective:

- Use Markov chain Monte Carlo (MCMC) methods to draw from complex distributions.

Implementation:

1. Gibbs sampling.
2. Metropolis-Hastings algorithm.

Ideas behind MCMC

- Goal: draw from complex target distribution
- Idea: target distribution is the stationary distribution of the Markov Chain
- Why it what is needed:
  1. once stationarity is reached: draw using transition matrix
  2. state is current draw
  3. by LLN fractions of draws $\approx$ target distribution
- How to construct transition matrix such that it has unique stationary distribution that equals target?

Transition matrix crafting

- Start with simple transition, e.g. symmetric random walk
- To achieve target as stationary: modify probabilities with accept/reject rule

Two variants of MCMC for today:

1. Gibbs: use marginal distribution, every draw is accepted
2. MH: use random walk, modify probabilities with accept/reject rule

Background

Motivation and intuition:

1. Draw from multivariate distributions from which direct sampling is difficult.
2. Construct conditional distributions from which sampling is easy.
3. Iteratively draw from conditional distributions.
4. Suppose we wish to sample $\theta_1, \theta_2 \sim p(\theta_1, \theta_2)$ but cannot do so directly.
5. However, we can sample $\theta_1 \sim p(\theta_1 \mid \theta_2)$ and $\theta_2 \sim p(\theta_2 \mid \theta_1)$.

Background

Algorithm:

1. Set $j = 0$.
2. Provide initial values $\left(\theta_1^{(0)}, \theta_2^{(0)}\right)$.
3. Set $j = j + 1$.
4. $\theta_1^{(j)} \sim p\left(\theta_1^{(j)} \mid \theta_2^{(j-1)}\right)$
5. $\theta_2^{(j)} \sim p\left(\theta_2^{(j)} \mid \theta_1^{(j)}\right)$
6. If j is less than the desired number of draws, return to step 3.

Exercise

Jupyter notebook:

1. Implement a Gibbs sampler to draw from a bivariate normal, i.e.

$$\begin{pmatrix} \theta_1 \\ \theta_2 \end{pmatrix} \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma}) \text{ with } \boldsymbol{\mu} = \mathbf{0}, \boldsymbol{\Sigma} = \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}, \rho = 0.8$$

2. Visually compare the empirical density of the draws to the theoretical density of the sampling distribution.

Note:

If:

$$\begin{pmatrix} \theta_1 \\ \theta_2 \end{pmatrix} \sim \mathcal{N}\left(\mathbf{0}, \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}\right)$$

then

$$\theta_1 \mid \theta_2 \sim \mathcal{N}\left(\rho\theta_2, 1-\rho^2\right)$$

$$\theta_2 \mid \theta_1 \sim \mathcal{N}\left(\rho\theta_1, 1-\rho^2\right)$$

Background

Motivation and intuition:

1. Gibbs sampling is only feasible, if the conditional distributions are known and if it is easy to sample from them.
2. The Metropolis-Hastings algorithm allows us to sample from any density $p(\theta)$, provided that we can evaluate $f(\theta)$, a density that is proportional to p.
3. At each iteration, we sample a new state from a candidate distribution, which depends on the current state.
4. With some probability depending on the value of the target density at the new and the current states, the new state is accepted.
5. As the algorithm proceeds, the sampled states approximate the desired density.

Background

Algorithm:

1. Set $j = 0$
2. Let θ denote the current state.
3. Let $p(\theta \mid x)$ denote the target density.
4. Let $J(\theta^* \mid \theta)$ denote the jumping distribution
5. Let ρ denote the step size.
6. Set $j = j + 1$.
7. Propose a new state $\theta^* \mid \theta \sim J(\theta)$.
8. Calculate $\alpha = \frac{p(\theta^*|x)}{p(\theta|x)} \cdot \frac{J(\theta|\theta^*)}{J(\theta^*|\theta)}$.
9. Draw $r \sim \mathrm{Uniform}(0, 1)$
10. If $r \leq \alpha$, accept the new state and set $\theta = \theta^*$. Otherwise, reject the new state.
11. If j is less than the desired number of draws, return to step 5.

Exercise

Jupyter notebook:

1. Simulate data from Cauchy distribution with location $\mu = 1$ and scale $\gamma = 1$
2. Implement a Metroplis-Hastings algorithms to infer the posterior distribution of the scale parameter of the Cauchy distribution.
3. Consider two possible jumping distributions:
   1. $\text{Normal}\left(\theta, \rho^2\right)$, where ρ is the step size.
   2. $\text{Lognormal}\left(\log(\theta) - 0.5\rho^2, \rho\right)$, where ρ is a distance parameter. Note that $\text{Lognormal}(\mu, \sigma)$ denotes a lognormal distribution with location μ and scale σ.
4. Evaluate the performance of the algorithms for different parametrisations of the jumping distributions. Compute the potential scale reduction factors $\hat{R}$

Exercise

Suppose:

- $x_i \sim \mathcal{C}(\mu, \gamma)$ for $i \in \{1, \ldots, N\}$
- $\gamma \sim \mathrm{Gamma}\,(\alpha_0, \beta_0)$ with $\alpha_0 = \beta_0 = 0.001$.

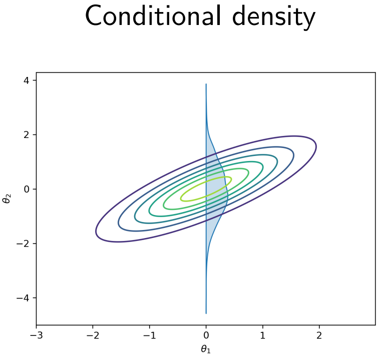
Then

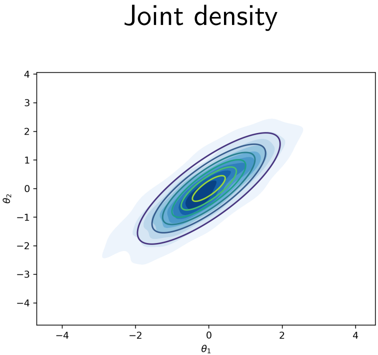
$$p\,(\gamma \mid x, \mu, \alpha_0, \beta_0) \propto \left(\prod_i p\,(x_i \mid \mu, \gamma)\right) p\,(\gamma \mid \alpha_0, \beta_0)$$

Note:

- PDF of Cauchy distribution: $f(x \mid \mu, \gamma) = \frac{1}{\pi\gamma\left[1+\left(\frac{x-\mu}{\gamma}\right)^2\right]}$
- Quantile function of Cauchy distribution: $Q = \mu + \gamma \cdot \tan\left[\pi\left(F - \frac{1}{2}\right)\right]$

Gibbs


Conditional density
θ_2
θ_1
4
2
0
−2
−4
−3
−2
−1
0
1
2


Joint density
θ_2
θ_1
4
3
2
1
0
−1
−2
−3
−4
−4
−2
0
2
4

Metropolis Hastings

Performance of the Normal jumping distribution for the different parameterizations:

ρ	Post. mean	Post. std.	Acceptance r	$\hat{R}$
0.00001	0.100627	0.000265	0.967	2.050453
0.00010	0.129414	0.005499	0.830	2.831171
0.00100	0.600619	0.070020	0.866	3.087170
0.01000	1.038503	0.043489	0.928	1.002975
0.10000	1.028516	0.045275	0.454	1.001987
0.15000	1.030084	0.044200	0.349	0.999190
0.20000	1.024537	0.043835	0.256	1.003908
0.25000	1.033822	0.042426	0.222	1.001908
0.50000	1.033683	0.047204	0.106	1.036941
1.00000	1.037752	0.038093	0.044	1.002151

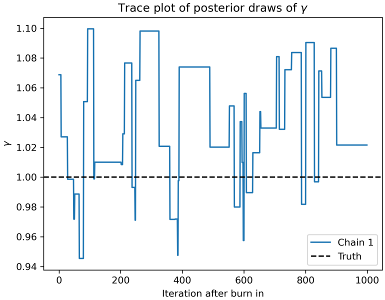
Metropolis Hastings

Performance of the LogNormal jumping distribution for the different parameterizations:

ρ	Post. mean	Post. std.	Acceptance r	$\hat{R}$
0.00001	0.100028	0.000015	0.999	1.673674
0.00010	0.100360	0.000130	0.974	2.598924
0.00100	0.137328	0.007989	0.764	2.953026
0.01000	1.052107	0.034412	0.921	1.006447
0.10000	1.032878	0.044744	0.433	1.001476
0.15000	1.024443	0.048473	0.317	1.002858
0.20000	1.028205	0.043205	0.258	1.012161
0.25000	1.025558	0.043160	0.205	1.026552
0.50000	1.034906	0.044986	0.107	1.010036
1.00000	1.033840	0.052076	0.046	1.003176

Metropolis Hastings

Normal jumping distribution



LogNormal jumping distribution