Sampling distributions with intractable normalizing constants using a Markov Chain Monte Carlo method *

Yice Luo, Tian Qin, Amir Kazemi

December 15, 2018

Contents

*Based on the work by Moller et al. (2006)

1 Goal

Proposition 1. *Consider the problem of Metropolis-Hastings sampling from*

$$\pi(\theta|y) \propto \pi(\theta)\pi(y|\theta) \tag{1}$$

where the likelihood is defined by

$$\pi(y|\theta) = q_\theta(y)/Z_\theta \tag{2}$$

with Z_θ as a normalizing constant. The Metropolis-Hastings ratio is

$$H(\theta'|\theta) = \frac{\pi(\theta')q_{\theta'}(y)p(\theta|\theta')}{\pi(\theta)q_\theta(y)p(\theta'|\theta)}\frac{Z_\theta}{Z_{\theta'}} \tag{3}$$

where $p(\theta'|\theta)$ denotes the proposal distribution.

Proof.

$$\begin{aligned} H(\theta'|\theta) &= \frac{\pi(\theta'|y)p(\theta|\theta')}{\pi(\theta|y)p(\theta'|\theta)} \\ &= \frac{\pi(\theta')\pi(y|\theta')p(\theta|\theta')}{\pi(\theta)\pi(y|\theta)p(\theta'|\theta)} \\ &= \frac{\pi(\theta')\left(q_{\theta'}(y)/Z_{\theta'}\right)p(\theta|\theta')}{\pi(\theta)\left(q_\theta(y)/Z_\theta\right)p(\theta'|\theta)} \\ &= \frac{\pi(\theta')q_{\theta'}(y)p(\theta|\theta')}{\pi(\theta)q_\theta(y)p(\theta'|\theta)}\frac{Z_\theta}{Z_{\theta'}} \end{aligned} \tag{4}$$

□

As Eq. (3) shows, the Metropolis-Hastings ratio depends on the unknowen fraction $Z_\theta/Z_{\theta'}$ which is usually approximated. The method in this report avoids apporixmation to $Z_\theta/Z_{\theta'}$ by considering an auxiliary variable which cancels $Z_\theta/Z_{\theta'}$.

2 Method

Proposition 2. *Let the auxiliary variable x be defined on the state space of y. The Metropolis-Hastings ratio for sampling from the posterior $\pi(\theta, x|y)$ becomes*

$$H(\theta', x'|\theta, x) = \frac{f(x'|\theta', y)\pi(\theta')q_{\theta'}(y)q_\theta(x)p(\theta|\theta')}{f(x|\theta, y)\pi(\theta)q_\theta(y)q_{\theta'}(x')p(\theta'|\theta)} \tag{5}$$

provided that

- *$p(x'|\theta', \theta, x) = \pi(x'|\theta')$, and*
- *$p(\theta'|\theta, x) = p(\theta'|\theta)$*

where $f(x|\theta, y)$ and $p(\theta'|\theta)$ are auxiliary and proposal densities, respectively. Then, (θ', x') is accepted with probability $\min(1, H(\theta', x'|\theta, x))$; otherwise, (θ, x) is accepted.

Proof.

$$\begin{aligned} H(\theta', x'|\theta, x) &= \frac{\pi(\theta', x'|y)p(\theta, x|\theta', x')}{\pi(\theta, x|y)p(\theta', x'|\theta, x)} \\ &= \frac{f(x'|\theta', y)\pi(\theta'|y) \times p(x|\theta, \theta', x')p(\theta|\theta', x')}{f(x|\theta, y)\pi(\theta|y) \times p(x'|\theta', \theta, x)p(\theta'|\theta, x)} \end{aligned} \tag{6}$$

The assumptions $p(x'|\theta', \theta, x) = \pi(x'|\theta')$ and $p(\theta'|\theta, x) = p(\theta'|\theta)$ give

$$H(\theta', x'|\theta, x) = \frac{f(x'|\theta', y)\pi(\theta'|y) \times \pi(x|\theta)p(\theta|\theta')}{f(x|\theta, y)\pi(\theta|y) \times \pi(x'|\theta')p(\theta'|\theta)} \tag{7}$$

and applying Eqs. (1-2) yield

$$\begin{aligned} H(\theta', x'|\theta, x) &= \frac{f(x'|\theta', y)\pi(\theta')\left(q_{\theta'}(y)/Z_{\theta'}\right) \times \left(q_\theta(x)/Z_\theta\right) p(\theta|\theta')}{f(x|\theta, y)\pi(\theta)\left(q_\theta(y)/Z_\theta\right) \times \left(q_{\theta'}(x')/Z_{\theta'}\right) p(\theta'|\theta)} \\ &= \frac{f(x'|\theta', y)\pi(\theta')q_{\theta'}(y)q_\theta(x)p(\theta|\theta')}{f(x|\theta, y)\pi(\theta)q_\theta(y)q_{\theta'}(x')p(\theta'|\theta)}. \end{aligned} \tag{8}$$

□

Corollary 1. *The ideal choice for auxiliary density $f(x|\theta, y)$ is*

$$f(x|\theta, y) = \frac{q_\theta(x)}{Z_\theta} \tag{9}$$

which gives $H(\theta', x'|\theta, x) = H(\theta'|\theta)$

Proof. Eq. (5) can be written as

$$H(\theta', x'|\theta, x) = \frac{\pi(\theta')q_{\theta'}(y)p(\theta|\theta')}{\pi(\theta)q_\theta(y)p(\theta'|\theta)} \times \frac{q_\theta(x)/f(x|\theta, y)}{q_{\theta'}(x')/f(x'|\theta', y)} \tag{10}$$

which may alternatively be obtained by subsituting normalizing constants in Eq. (3) as follows

$$\begin{aligned} Z_\theta &= q_\theta(x)/f(x|\theta, y) \\ Z'_\theta &= q_{\theta'}(x')/f(x'|\theta, y). \end{aligned} \tag{11}$$

In another view, Eq. (11) is one-sample importance sampling for

$$\begin{aligned} Z_\theta &= \mathrm{E}\left[q_\theta(x)\right] = \mathrm{E}_{g_\theta}\left[q_\theta(x)/g_\theta(x)\right] \\ Z'_\theta &= \mathrm{E}\left[q_{\theta'}(x')\right] = \mathrm{E}_{g_\theta}\left[q_{\theta'}(x')/g_{\theta'}(x')\right] \end{aligned} \tag{12}$$

where $g_\theta(x)$ and $g_{\theta'}(x')$ are instrumental densities for importance sampling. $\square$

Finding the ideal choice for $f(x|\theta, y)$ is impractical; therefore, if the posterior $\pi(x|\theta) = q_\theta(x)/Z_\theta$ does not strongly depend on θ, a simple choice for the auxiliary density is

$$f(x|\theta, y) = \frac{q_{\tilde{\theta}}(x)}{Z_{\tilde{\theta}}} \tag{13}$$

where $\tilde{\theta} = \tilde{\theta}(y)$ estimates θ in terms of y and the normalizing constants of the auxiliary density is canceled in Eq. (5) i.e.

$$H(\theta', x'|\theta, x) = \frac{q_{\tilde{\theta}}(x)\pi(\theta')q_{\theta'}(y)q_\theta(x)p(\theta|\theta')}{q_{\tilde{\theta}}(x')\pi(\theta)q_\theta(y)q_{\theta'}(x')p(\theta'|\theta)}. \tag{14}$$

3 Application: The Ising Model

A common example of densities with intractable normalizing constants is Ising model, which is a special case of autologistic model. The unnormalized probability measure of state configurations for the sites of a $m \times n$ lattice is obtained by

$$q_\theta(y) = \exp(\theta_0 V_0 + \theta_1 V_1) \tag{15}$$

where y is a $m \times n$ matrix representing the state realization of the sites with $y_{i,j} = \{-1, +1\}$, θ_0 and θ_1 are constants for external field and assoication states together, and

$$\begin{aligned} V_0 &= \sum_{i=1}^{m}\sum_{j=1}^{n} y_{i,j} \\ V_1 &= \sum_{i=1}^{m-1}\sum_{j=1}^{n} y_{i,j}y_{i+1,j} + \sum_{i=1}^{m}\sum_{j=1}^{n-1} y_{i,j}y_{i,j+1}. \end{aligned} \tag{16}$$

Therefore, $\theta_1 = 0$ means that $y_{i,j}$ do not assoicate together and are i.i.d.

4 Algorithm

Data:
$\boldsymbol{\theta}_0 = (\theta_0^{(0)}, \theta_1^{(0)})$: initial value for (θ_0, θ_1)
MSS: maximum sample size
i: iteration index
Result: $\boldsymbol{\theta}_i$ for $i \leq MSS$
while $\theta_0^{(0)} \notin [-1, 1]$ *or* $\theta_1^{(0)} \notin [0, 1]$ **do**
    Choose $\boldsymbol{\theta}_0 = (\theta_0^{(0)}, \theta_1^{(0)})$ s.t. $\theta_0^{(0)} \in [-1, 1]$ and $\theta_1^{(0)} \in [0, 1]$;
    $\boldsymbol{\theta}_0 \leftarrow$ MPLE of the given $\boldsymbol{\theta}_0$
end
Draw $\boldsymbol{x} \sim q_{\boldsymbol{\theta}}(\boldsymbol{x})$ by perfect sampling or Gibbs sampler
while $i < MSS$ **do**
    Draw $\theta_0^{\prime(i)} \sim \mathcal{N}(\theta_0^{(i)}, \sigma)$
    Draw $\theta_1^{\prime(i)} \sim \mathcal{N}(\theta_1^{(i)}, \sigma)$
    $\boldsymbol{\theta}_i' \leftarrow (\theta_0^{\prime(i)}, \theta_1^{\prime(i)})$
    Draw $\boldsymbol{x'} \sim q_{\boldsymbol{\theta}'}(\boldsymbol{x}')$ by perfect sampling or Gibbs sampler
    Draw $U \sim U[0, 1]$
    $H \leftarrow H(\boldsymbol{\theta}', \boldsymbol{x}' | \boldsymbol{\theta}_i, \boldsymbol{x})$ from Eq. (14)
    if $U \leq H$ **then**
        $(\boldsymbol{\theta}_{i+1}, \boldsymbol{x}) \leftarrow (\boldsymbol{\theta}', \boldsymbol{x}')$
    else
        $(\boldsymbol{\theta}_{i+1}, \boldsymbol{x}) \leftarrow (\boldsymbol{\theta}_i, \boldsymbol{x})$
    end
end

Algorithm 1: Sampling $\boldsymbol{\theta}$

.

	Case ID				
	case1	case2	case3	casePERFECT	caseGIBBS
Sampling method	perfect	perfect	perfect	perfect	Gibbs
Size of lattice	50	50	50	5	5
σ	0.01	0.005	0.005	0.005	0.005
$\theta_0^{(0)}$	0.2	0.1	0.2	0.2	0.2
$\theta_1^{(0)}$	0.1	0.3	0.1	0.1	0.1
MSS (iterations)	1e4	1e4	1e4	1e3	1e3

5 Results

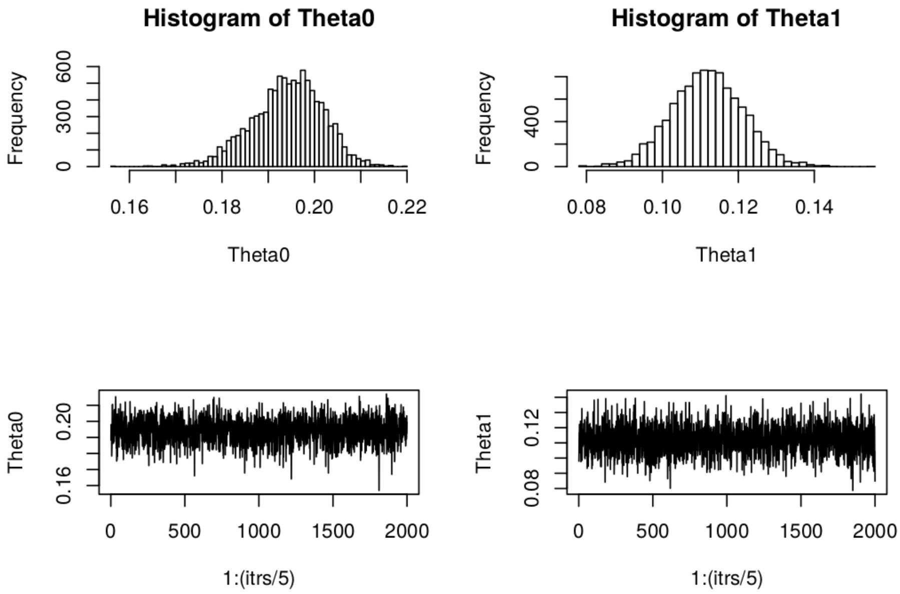

Figure 1: Case1

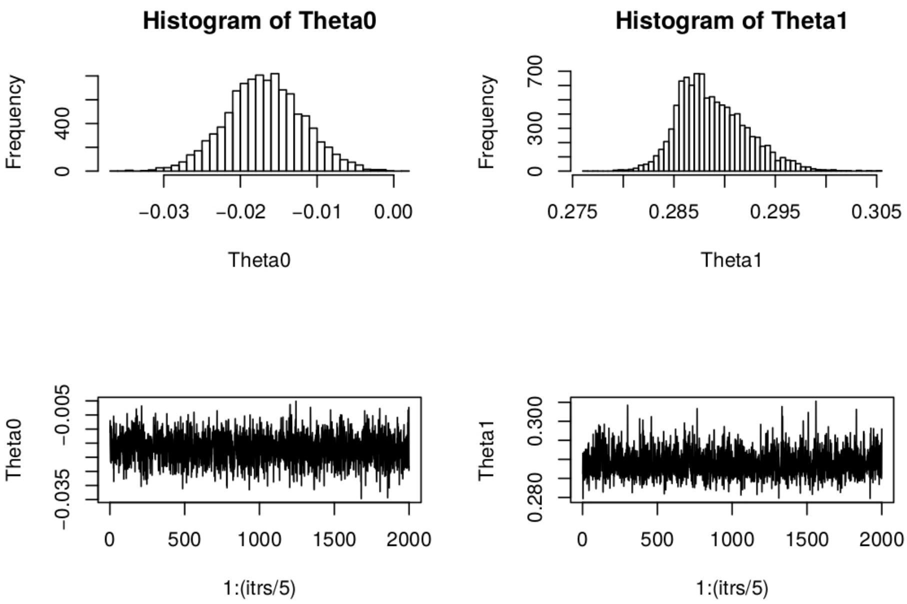

Figure 2: Case2

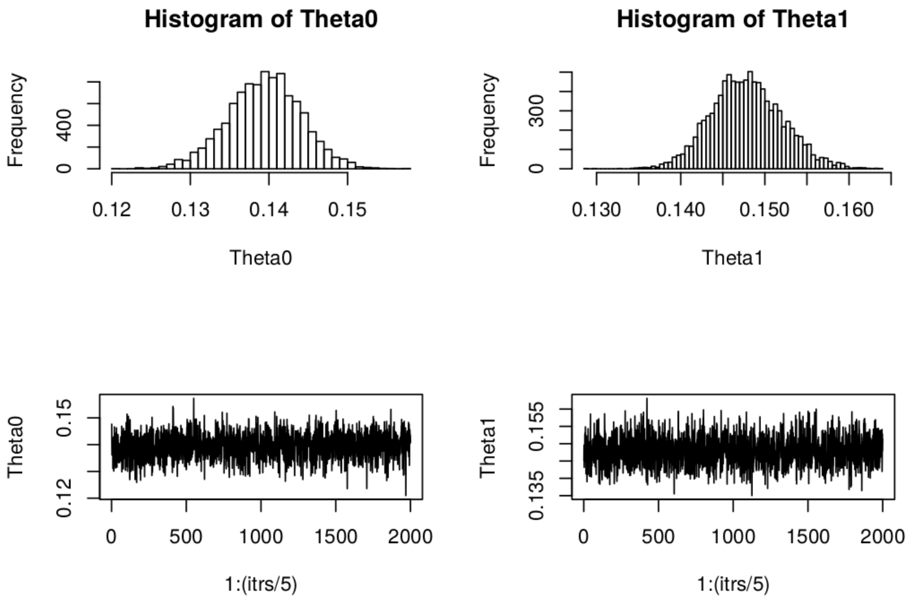

Figure 3: Case3

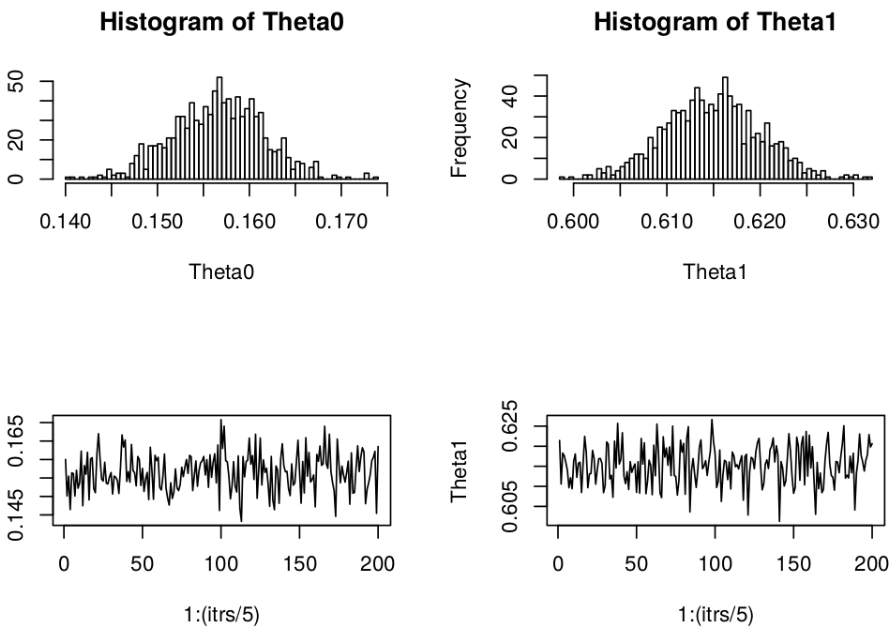

Figure 4: CasePERFECT

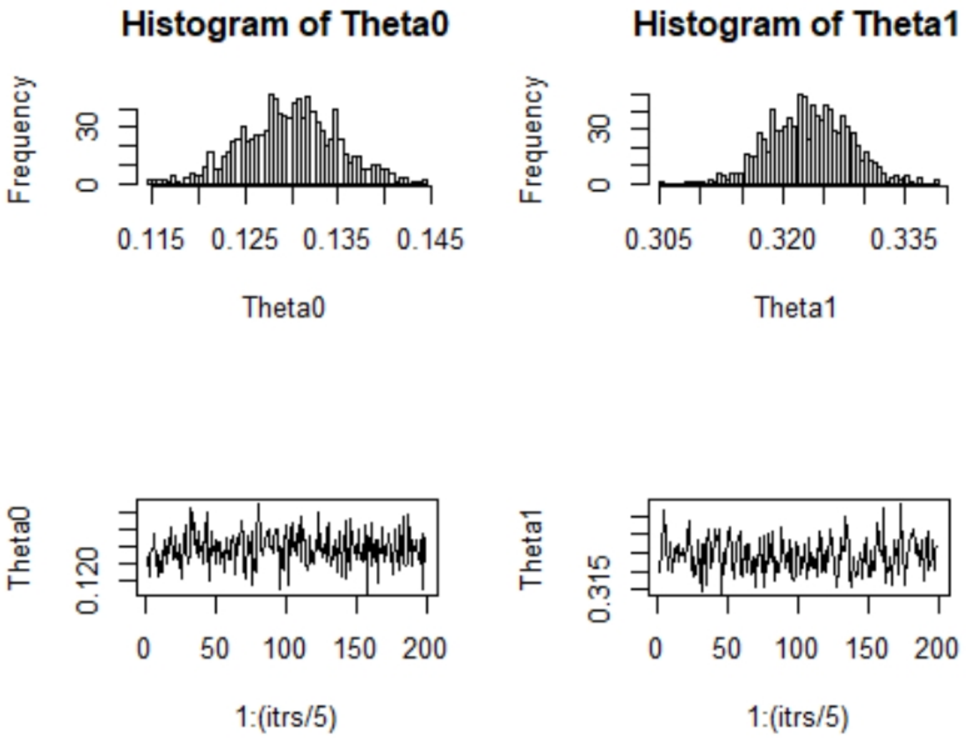

Figure 5: CaseGIBBS

6 References

Moller, J., Pettitt, A. N., Reeves, R., & Berthelsen, K. K. (2006). An efficient Markov chain Monte Carlo method for distributions with intractable normalising constants. Biometrika, 93(2), 451-458.

7 Appendix: Codes