Hamiltonian Monte Carlo

1 Problem Statement

We consider the stochastic differential equation

$$\mathrm{d}X_t = f(X_t;\theta)\mathrm{d}t + g(X_t;\theta)\mathrm{d}W_t \tag{1}$$

where $X = \{X_t\}_{t\geq 0} \subseteq \mathbb{R}$, $\theta \subseteq \mathbb{R}^k$ and $W = \{W_t\}_{t\geq 0}$ is the standard Brownian motion. The drift function $f(X_t;\theta)$ and the diffusion function $g(X_t;\theta)$ are unknown functions which we will try to infer. Initially we assume that the functions are known in a parametric sense and we infer the parameters using the approach as explained in the following sections.

The discretized form of the equation can be written using Euler-Maruyama approximation as

$$X_{n+1} = X_n + f(X_n;\theta)\Delta t + g(X_n;\theta)\mathcal{N}(0,\Delta t) \tag{2}$$

where $n = 0,\ldots,N$ where the time discretization is $t_0,\ldots,t_N$. The transition density of the process can be denoted by $P_{n+1}(x_{n+1}|X_n = x_n)$ evolving from X_{n+1} to X_n.

2 Methods - Transition density and gradients

Likelihood $L(\theta) =$
Gradient $\frac{\partial}{\partial\theta}L(\theta) =$

3 HMC

The classic Metropolis-Hastings algorithm for bayesian inference has some In Hamiltonian mechanics, the time evolution of a system is governed by the Hamilton's equations:

$$\frac{dp}{dt} = -\frac{dH}{dq}$$
$$\frac{dq}{dt} = \frac{dH}{dp}$$

We represent the posterior density as $P(\theta|X)$ where θ is the parameters vector and X is the data. Then the total energy of the system can be written as the summation of kinetic energy (K) and the potential energy (V) and $H = V + K$.

$$\begin{aligned} P(\theta|X) &= \frac{1}{Z}\exp\bigg(-\frac{H(\theta,\phi)}{T}\bigg) \\ &= \frac{1}{Z}\exp\bigg(-\frac{K(\theta,\phi)}{T} - \frac{V(\theta,\phi)}{T}\bigg) \\ log(ZP(\theta|X)) &= -\frac{V}{T} \end{aligned}$$

If we consider $T = 1$ and mass m as the variance, the leapfrog scheme can be written as,

$$\begin{aligned} V &= -\log(P(\theta|X)) \\ K &= \frac{\phi^2}{2m} \end{aligned}$$

To compute the joint density, $P(\theta, \phi)$

$$
\begin{aligned}
P(\theta, \phi|X) &= \frac{1}{Z} \exp\Big(- \frac{\phi^2}{2m} + \log P(\theta|X) \Big) \\
&= \underbrace{\frac{1}{Z}}_{\sqrt{2\pi m}} \underbrace{e^{-\frac{\phi^2}{2m}}}_{\text{Gaussian}} \underbrace{P(\theta|X)}_{\text{Posterior}}
\end{aligned}
$$

Leapfrog scheme:

$$
\begin{aligned}
p_{n+\frac{1}{2}} &= p_n - V'(q_n)\frac{h}{2} \\
q_{n+1} &= q_n + \frac{p_{n+\frac{1}{2}}}{m} h \\
p_{n+1} &= p_{n+\frac{1}{2}} - V'(q_{n+1})\frac{h}{2}
\end{aligned}
$$

4 Hamiltonian Monte Carlo algorithm

1. Draw ϕ^n from a Gaussian distribution with mean 0 and variance m
2. Calculate an intermediate step, $\phi^{n+\frac{1}{2}} = \phi^n + \frac{d}{d\theta} \log(P(\theta^n|X))\frac{h}{2}$
3. The proposal θ^* is calculated as, $\theta^{n+1} = \theta^n + \frac{\phi^{n+\frac{1}{2}} h}{m}$
4. The proposal ϕ^* is calculated as, $\phi^n = \phi^{n+\frac{1}{2}} + \frac{d}{d\theta} \log(P(\theta^{n+1}|X))\frac{h}{2}$
5. Compute the ratio, $\rho = \frac{P(\theta^*, \phi^*|X)}{P(\theta^n, \phi^n|X)} = \frac{P(\theta^*|X)}{P(\theta^n|X)} \frac{P(\phi^*|X)}{P(\phi^n|X)}$
6. If $r > \mathcal{U}(0, 1)$, accept $\theta^{n+1} = \theta^*$, else reject, $\theta^{n+1} = \theta^n$

Comments

1. The error in the value of the Hamiltonian determines the rejection rate
2. The error usually does not increase with the number of leapfrog steps taken, provided that the stepsize is small enough that the dynamics is stable

5 Details of the code

The code is written in 2 parts. The main interface is through R which takes a RData file with the time series data with each row being a single time series with values for regular sized intervals $\{t\}_{n=0}^{n=N}$ such that $t_n = nh$. The R code calls a C++ function, Rgdtq, which uses the DTQ method to evaluate the objective function and the gradient with respect to all parameters for the specified SDE.

6 Results

In this section we present the results for a specific example, the Ornstein Uhlenback process which has linear drift and diffusion terms.

$$
dX_t = \theta_1(\theta_2 - X_t)dt + \theta_3^2\, dW_t \tag{3}
$$

The diffusion parameter is assumed to be non-negative and thus we consider a squared term as the diffusion parameter.

Table 1: Table with all the results (true = (0.8, 0.9, 0.7), fakedatah = $1e-6$, bigt = 25, h = 0.05, k = $0.05^{(0.85)}$, bigm = $\pi/k^{(1.5)}$, total steps = 500 with burnin = 100)

	Init	ntrials	epsilon	steps (L)	mass	iter	Accept%	RMSE	time
7 (varying θ_1)	(0.5, 0.5, 0.5)	50	0.001	20	1	500	79.4%		
11 (constant θ_1)	(0.8, 0.5, 0.5)	100	0.005	20	1	500	72.83%	0.0093	
10 (constant θ_1)	(0.8, 0.5, 0.5)	100	0.001	20	1	500	66.66%	0.0094	
8 (constant θ_1)	(0.8, 0.5, 0.5)	100	0.001	10	1	500	76.16%	0.0105	
9 (constant θ_1)	(0.8, 0.5, 0.5)	100	0.001	5	1	500	79.27%	0.0142	
12 (constant θ_1)	(0.8, 0.5, 0.5)	100	0.001	20	1	2000		78.19%	
13 (varying θ_1)	(0.5, 0.5, 0.5)	100	0.001	20	1	500			

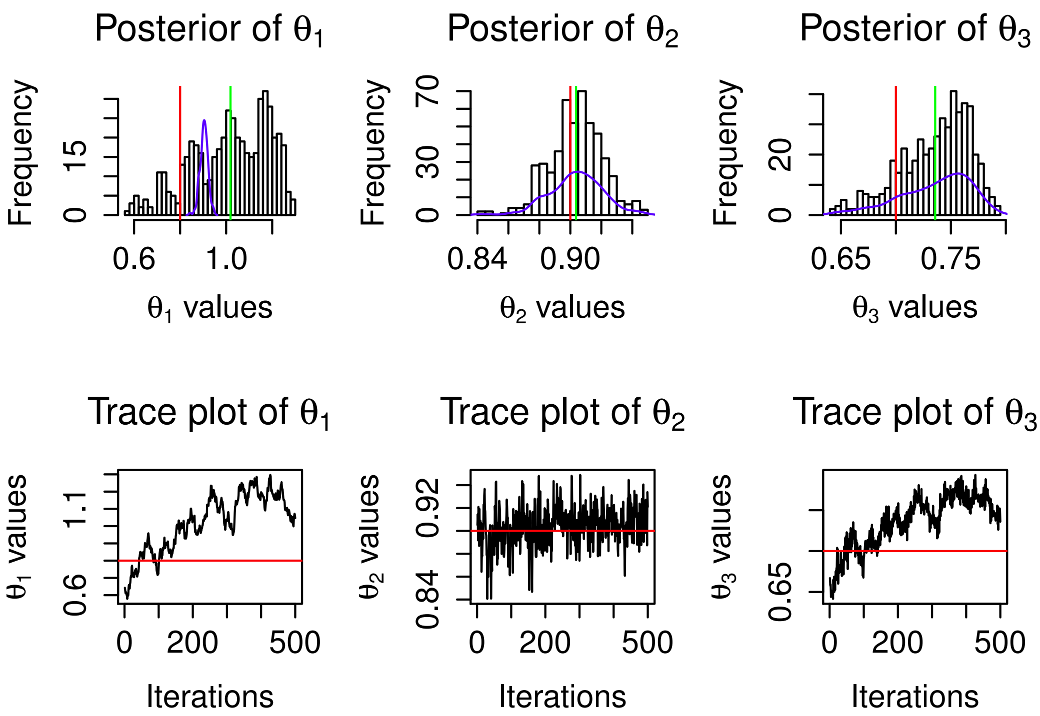


Figure 1: HMC plot 7 with positive θ_3 and varying θ_1

Listing 1: R output for plot 7 using the library CODA with varying θ_1 and 20 steps of leapfrog with 0.001 ϵ

```
Iterations = 1:500
Thinning interval = 1
Number of chains = 1
Sample size per chain = 500

1. Empirical mean and standard deviation for each variable, plus standard error of the mean:

       Mean       SD  Naive SE Time-series SE
[1,] 1.0188 0.16728 0.0074810       0.086829
[2,] 0.9037 0.01724 0.0007708       0.001382
[3,] 0.7357 0.03196 0.0014293       0.015289

2. Quantiles for each variable:
       2.5%    25%    50%    75%  97.5%
var1 0.6463 0.8877 1.0351 1.1636 1.2553
var2 0.8696 0.8930 0.9047 0.9150 0.9383
var3 0.6621 0.7151 0.7420 0.7606 0.7829
```

For better inference with HMC, we consider the first parameter to be a constant value of $\theta_1 = 0.8$

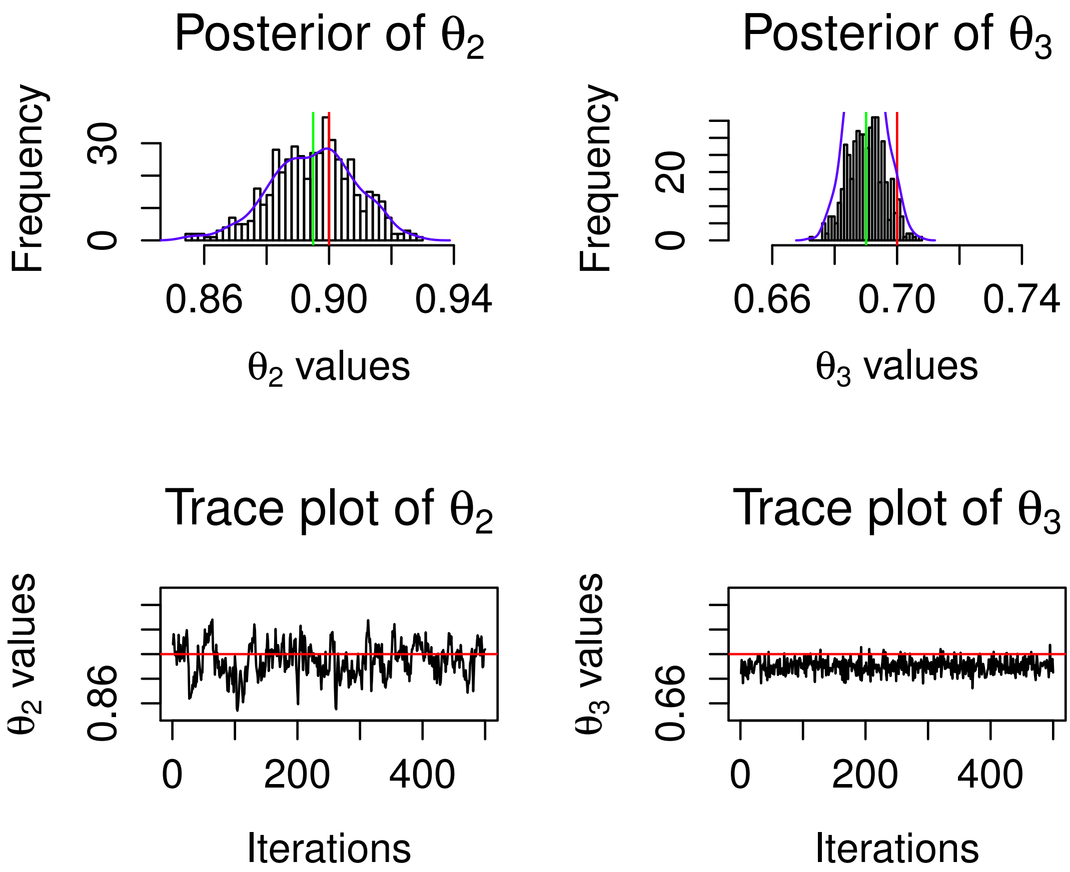


Figure 2: Trace plots, density and histograms for HMC 8 with constant $\theta_1 = 0.8$ and positive θ_3 with 5 leapfrog steps

Listing 2: R output for plot 8 using the library CODA with 10 steps and 0.001 ϵ

```
Iterations = 1:500
Thinning interval = 1
Number of chains = 1
Sample size per chain = 500

1. Empirical mean and standard deviation for each variable,
   plus standard error of the mean:

        Mean        SD  Naive SE Time-series SE
[1,] 0.8949 0.013593 0.0006079      0.0015178
[2,] 0.6901 0.005828 0.0002606      0.0002158

2. Quantiles for each variable:

       2.5%    25%    50%    75%  97.5%
var1 0.8677 0.8858 0.8957 0.9040 0.9187
var2 0.6789 0.6860 0.6901 0.6939 0.7013
```

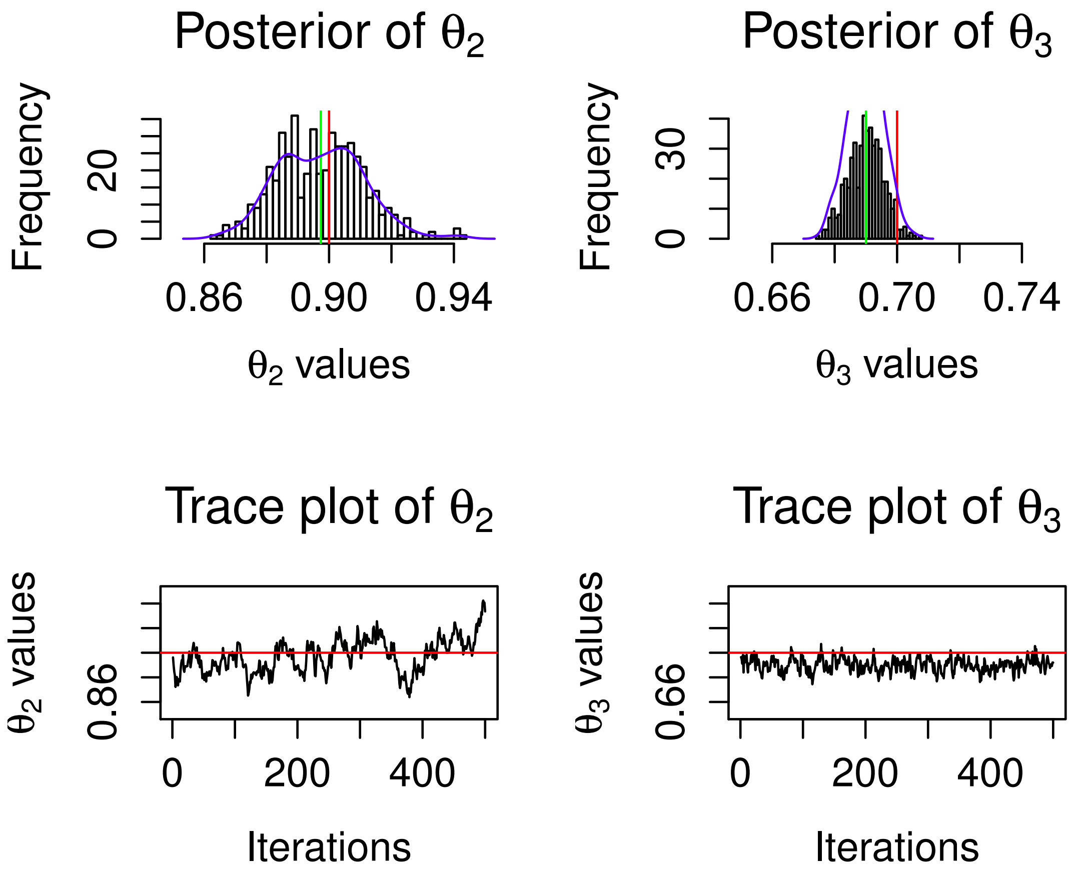


Figure 3: Trace plots, density and histograms for HMC 9 with constant $\theta_1 = 0.8$ and positive θ_3 with 10 leapfrog steps

Listing 3: R output for plot 9 using the library CODA with 5 steps and 0.001 ϵ

```
Iterations = 1:500
Thinning interval = 1
Number of chains = 1
Sample size per chain = 500

1. Empirical mean and standard deviation for each variable,
   plus standard error of the mean:

        Mean        SD  Naive SE Time-series SE
[1,] 0.8949 0.013593 0.0006079      0.0015178
[2,] 0.6901 0.005828 0.0002606      0.0002158

2. Quantiles for each variable:

       2.5%    25%    50%    75%  97.5%
var1 0.8677 0.8858 0.8957 0.9040 0.9187
var2 0.6789 0.6860 0.6901 0.6939 0.7013
```

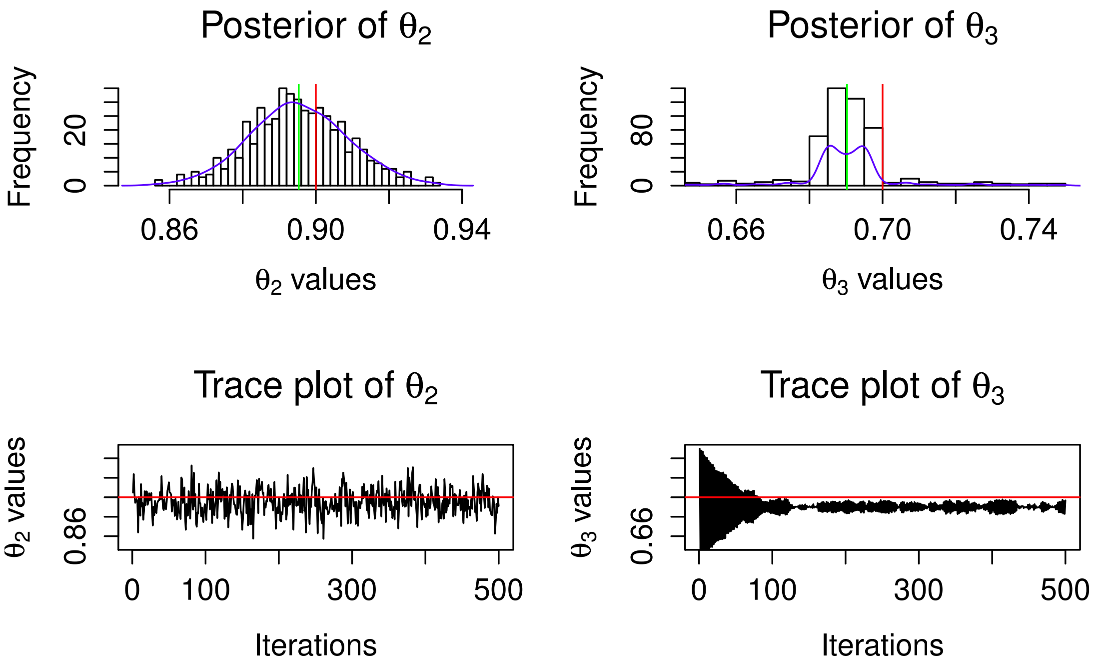


Figure 4: Trace plots, density and histograms for HMC 10 with constant $\theta_1 = 0.8$ and positive θ_3 with 20 leapfrog steps and $\epsilon = 0.001$

Listing 4: R output for plot 10 using the library CODA with 20 steps and 0.001 ϵ

```
Iterations = 1:500
Thinning interval = 1
Number of chains = 1
Sample size per chain = 500

1. Empirical mean and standard deviation for each variable,
   plus standard error of the mean:

        Mean       SD  Naive SE Time-series SE
[1,] 0.8953 0.01327 0.0005936      0.0006928
[2,] 0.6903 0.01364 0.0006101      0.0001241

2. Quantiles for each variable:

        2.5%    25%    50%    75%  97.5%
var1 0.8695 0.8864 0.8951 0.9041 0.9218
var2 0.6561 0.6852 0.6900 0.6950 0.7264
```

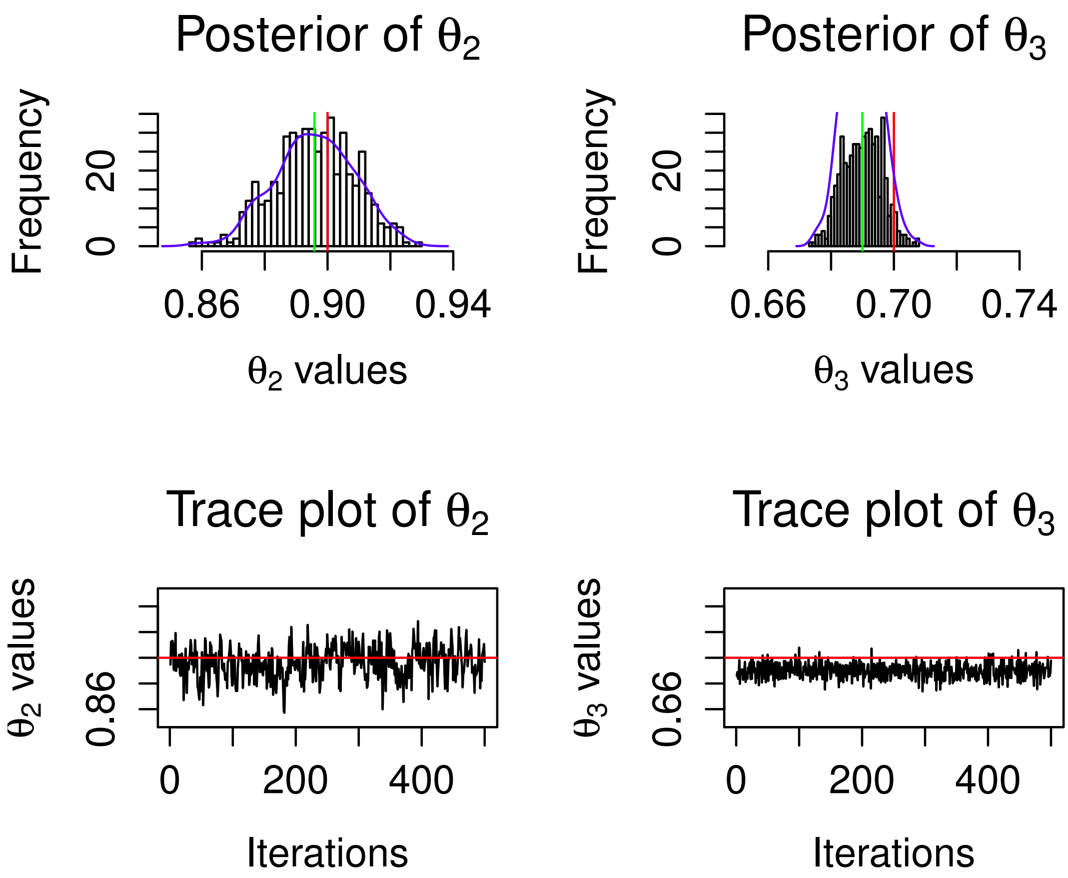
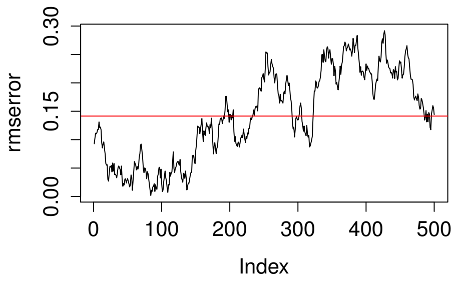


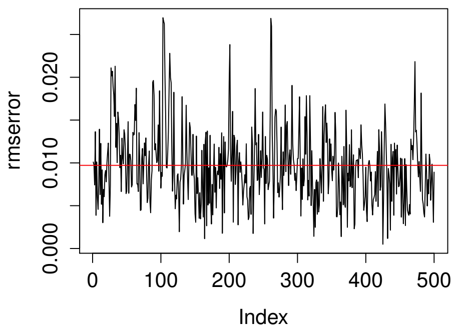
Figure 5: Trace plots, density and histograms for HMC 11 with constant $\theta_1 = 0.8$ and positive θ_3 with 20 leapfrog steps

Listing 5: R output for plot 11 using the library CODA with 20 steps and 0.005 ϵ

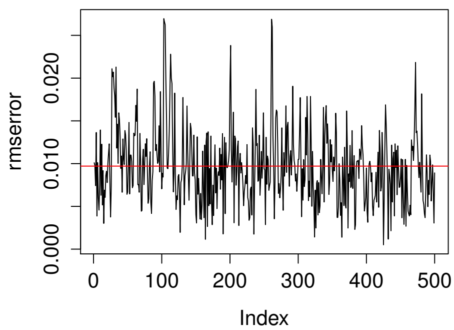
```
Iterations = 1:500
Thinning interval = 1
Number of chains = 1
Sample size per chain = 500

1. Empirical mean and standard deviation for each variable,
   plus standard error of the mean:

        Mean       SD  Naive SE Time-series SE
[1,] 0.8958 0.012625 0.0005646      0.0008973
[2,] 0.6900 0.006193 0.0002770      0.0002030

2. Quantiles for each variable:

       2.5%    25%    50%    75%  97.5%
var1 0.8727 0.8877 0.8960 0.9045 0.9198
var2 0.6783 0.6853 0.6901 0.6943 0.7013
```



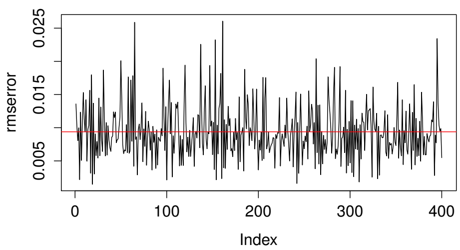
(a) HMC Plot 7



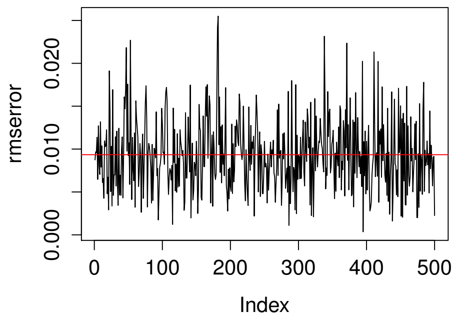
(b) HMC Plot 8



(c) HMC Plot 9



(d) HMC Plot 10



(e) HMC Plot 11

Figure 6: Error plot after removing burnin period for HMC plot

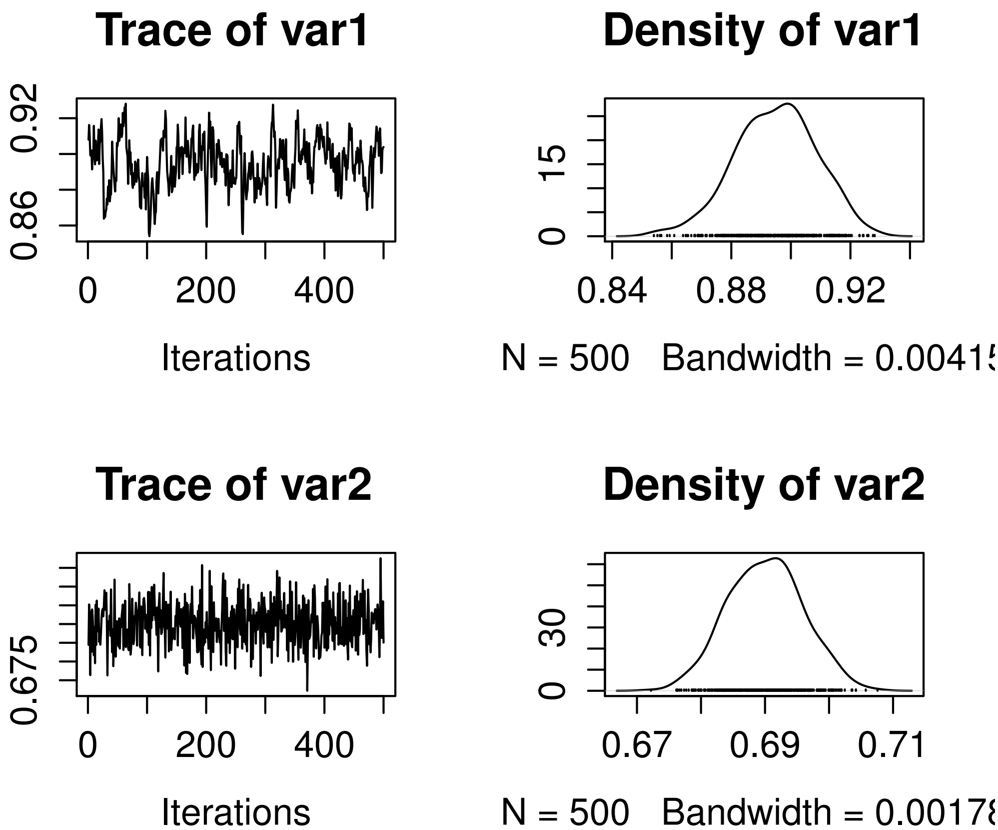


Figure 7: HMC plot 8 summary

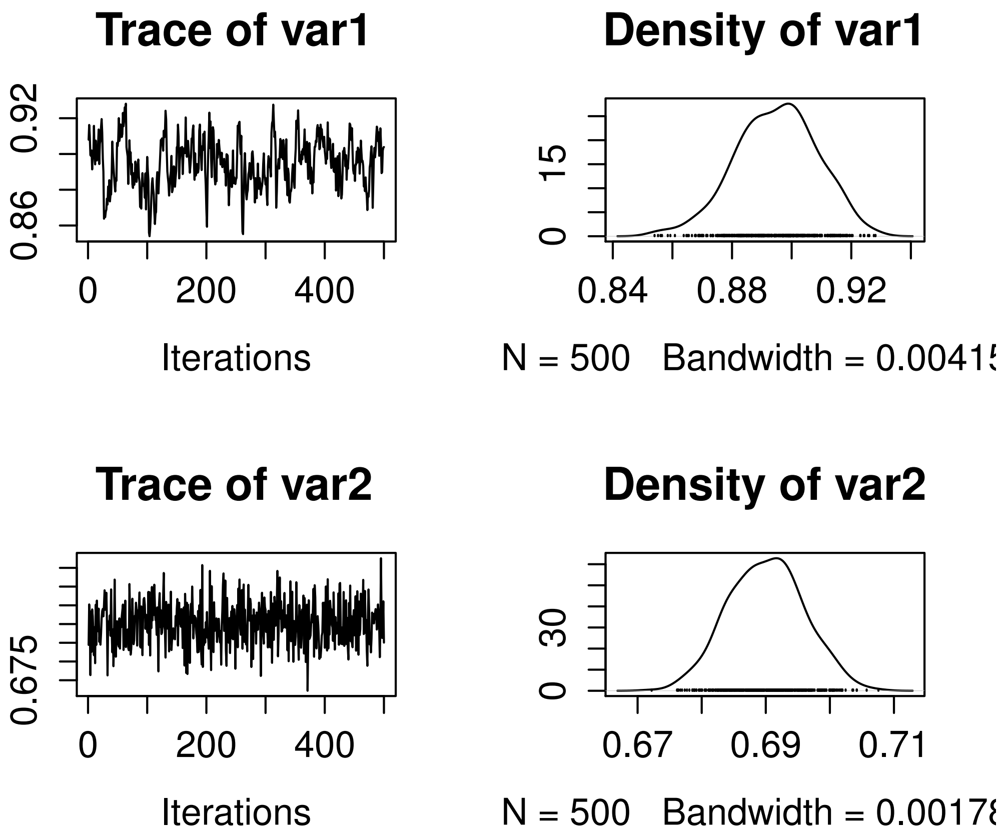


Figure 8: HMC plot 9 summary