07/04/2022. LECTURE 18: NOTES MADE ON THE VISUALISER.

Example: sampling from $N(0,1)$ using the proposal $N(\theta^{(t-1)}, \sigma^2)$ when the chain is at $\theta^{(t-1)}$ and σ^2 is known.

① Choose $\sigma^2 = 1$. Start point $\theta^{(0)} = 1$.

Figure 3.1 metropolis.update(sig.q = 1)

* See plots at the end *

Figure 3.2 metropolis(sig.q = 1, n = 100)

Figure 3.3 metropolis(sig.q = 1, n = 5000) Acceptance rate ≈ 70%

② Choose $\sigma^2 = 0.36$. Start point $\theta^{(0)} = 1$.

Figure 3.4 metropolis.update(sig.q = 0.6)

* See plots at the end *

Figure 3.5 metropolis(sig.q = 0.6, n = 100)

Figure 3.6 metropolis(sig.q = 0.6, n = 5000)

When $\sigma^2 = 0.36$ we make smaller jumps, leading to a higher autocorrelation (acceptance rate ≈ 76%). Reduce mixing: move around less.

③ Extreme case: tiny acceptance rate (huge variance)

metropolis(sig.q = 400, n = 100)

metropolis(sig.q = 400, n = 500 000)

④ Extreme case: slow mixing, bad start point

metropolis(sig.q = 0.001, start = -5, n = 100)

metropolis(sig.q = 0.001, start = -5, n = 1000 000)

Note: proposal determines how we obtain θ^*. Choice of σ^2 will determine the acceptance rate ("tuning parameter")

σ^2 small: high prob of acceptance, make small moves
(large auto correlation)

σ^2 large: small prob of acceptance, make large moves
(jagged; small acceptance rate)

[0.234 asymptotic optimal acceptance rate]

3.3.3 THE GIBBS SAMPLER

The Gibbs sampler aims to make sampling from a high-dimensional distribution more tractable by sampling from a collection of smaller dimensional distribution

It's an MCMC scheme where the transition kernel is formed by the full conditionals. Assume that the distribution of interest is $\pi(\theta)$ where $\theta = (\theta_1, \dots, \theta_d)$.

1. Typically we want $\pi(\theta) = f(\theta | x)$
2. Components θ_i can be scalar, vectors, matrices. We'll regard them as scalars.

Let θ_{-i} denote $\theta / \theta_i = (\theta_1, \dots, \theta_{i-1}, \theta_{i+1}, \dots, \theta_d)$ and suppose that the full conditionals $\pi_i(\theta_i) = \pi(\theta_i | \theta_{-i})$ $i = 1, \dots, d$ are available and can be sampled from.

[If $\pi(\theta) = f(\theta | x)$ then $\pi_i(\theta_i) = f(\theta_i | \theta_{-i}, x)$
$= f(\theta_i | \theta_1, \dots, \theta_{i-1}, \theta_{i+1}, \dots, \theta_d, x)$]

The Gibbs sampler algorithm.

1). Choose an arbitrary starting point $\theta^{(0)} = (\theta_1^{(0)}, \dots, \theta_d^{(0)})$ for which $\pi(\theta^{(0)}) > 0$.

2) At time t, $[\theta^{(t-1)} = (\theta_1^{(t-1)}, \theta_2^{(t-1)}, \dots, \theta_d^{(t-1)})]$

- Sample $\theta_1^{(t)}$ from conditional distribution

$$\pi(\theta_1 \mid \theta_2 = \theta_2^{(t-1)}, \dots, \theta_d = \theta_d^{(t-1)})$$
$$= \pi(\theta_1 \mid \theta_2^{(t-1)}, \dots, \theta_d^{(t-1)})$$

- Sample $\theta_2^{(t)}$ from conditional distribution

$$\pi(\theta_2 \mid \theta_1^{(t)}, \theta_3^{(t-1)}, \dots, \theta_d^{(t-1)})$$

$\vdots$

- Sample $\theta_p^{(t)}$ from conditional distribution

$$\pi(\theta_p \mid \underbrace{\theta_1^{(t)}, \dots, \theta_{p-1}^{(t)}, \theta_{p+1}^{(t-1)}, \dots, \theta_d^{(t-1)}}_{\text{always use latest value of each } \theta_i})$$

$\vdots$

- Sample $\theta_d^{(t)}$ from conditional distribution

$$\pi(\theta_d \mid \theta_1^{(t)}, \dots, \theta_{d-1}^{(t)})$$

3). Repeat step 2).

The algorithm is run until convergence is reached. When this occurs, the resulting value (and all subsequent values) $\theta^{(j)}$ is a realisation from $\pi(\theta)$.

Note:

$$\pi(\theta_i \mid \theta_{-i}) = \frac{\pi(\theta_i, \theta_{-i})}{\pi(\theta_{-i})} = \frac{\pi(\theta_1, \dots, \theta_d)}{\pi(\theta_{-i})}$$

function of θ_i

does not depend on θ_i

$$\propto \pi(\theta_1, \dots, \theta_d)$$

with respect to θ_i

[Posterior distribution: $f(\theta_i \mid \theta_{-i}, x) \propto f(\theta \mid x)$ as a function of θ_i]

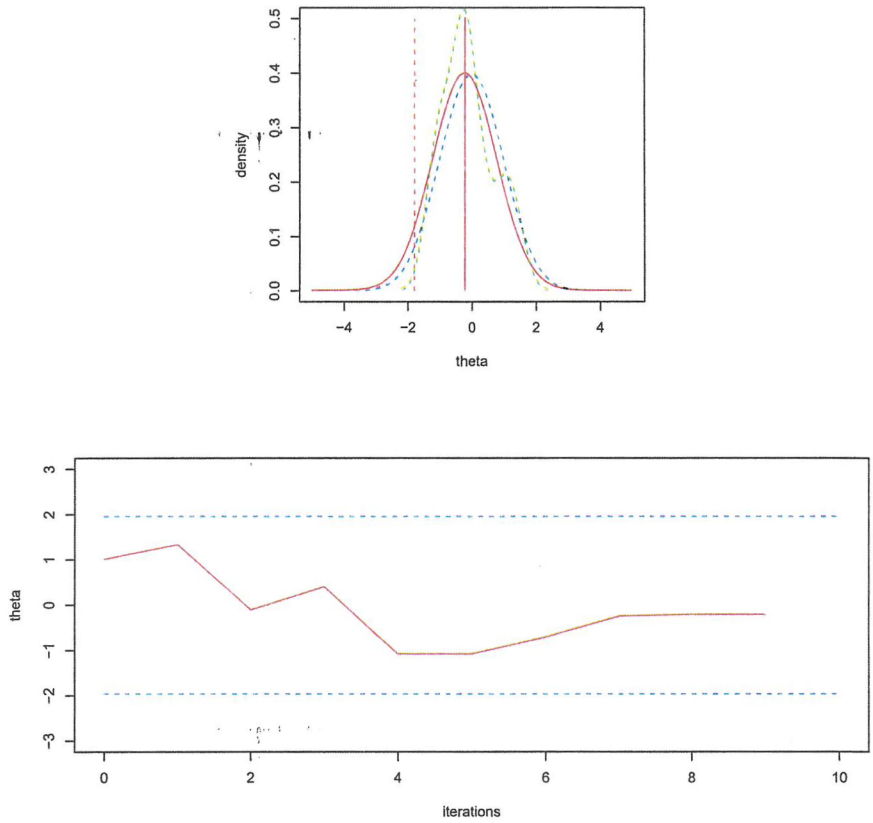

Figure 1: Nine iterations from a Metropolis-Hastings sampler for $\theta \mid x \sim N(0,1)$ where the proposal distribution is normal with mean θ and chosen variance $\sigma^2 = 1$ and the starting point is $\theta^{(0)} = 1$. The top plot shows the target distribution, $N(0,1)$, the proposal distribution for $\theta^{(9)}$ which is $N(-0.216, 1)$, the proposal $\theta^* = -1.779$ and the current observed density. The move is rejected; the bottom plot is the trace plot of $\theta^{(t)}$.

t	1	2	3	4	5	6	7	8	9
θ^*	1.319	−0.118	0.392	−1.089	−1.947	−0.716	−0.251	−0.216	−1.779
Accept/Reject?	Accept	Accept	Accept	Accept	Reject	Accept	Accept	Accept	Reject
$\theta^{(t)}$	1.319	−0.118	0.392	−1.089	−1.089	−0.716	−0.251	−0.216	−0.216

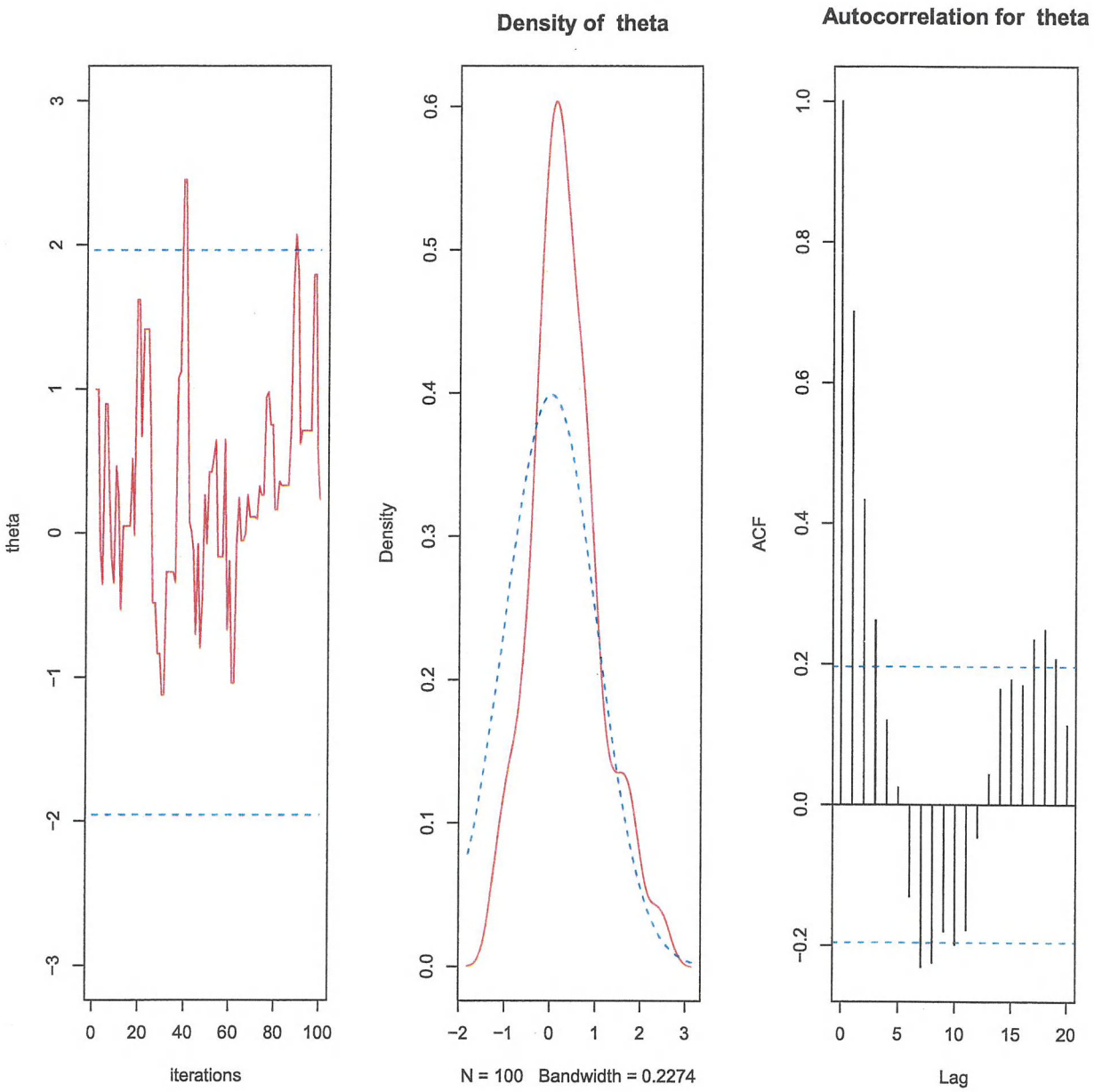

Figure 2: 100 iterations from a Metropolis-Hastings sampler for $\theta \,|\, x \sim N(0,1)$ where the proposal distribution is normal with mean θ and chosen variance $\sigma^2 = 1$ and the starting point is $\theta^{(0)} = 1$. The first plot shows the trace plot of $\theta^{(t)}$, the second the observed density against the target density and the third the autocorrelation plot. The sample mean is 0.3435006 and the sample variance is 0.5720027. 3% of points were observed to be greater than 1.96 and the acceptance rate is 0.66.

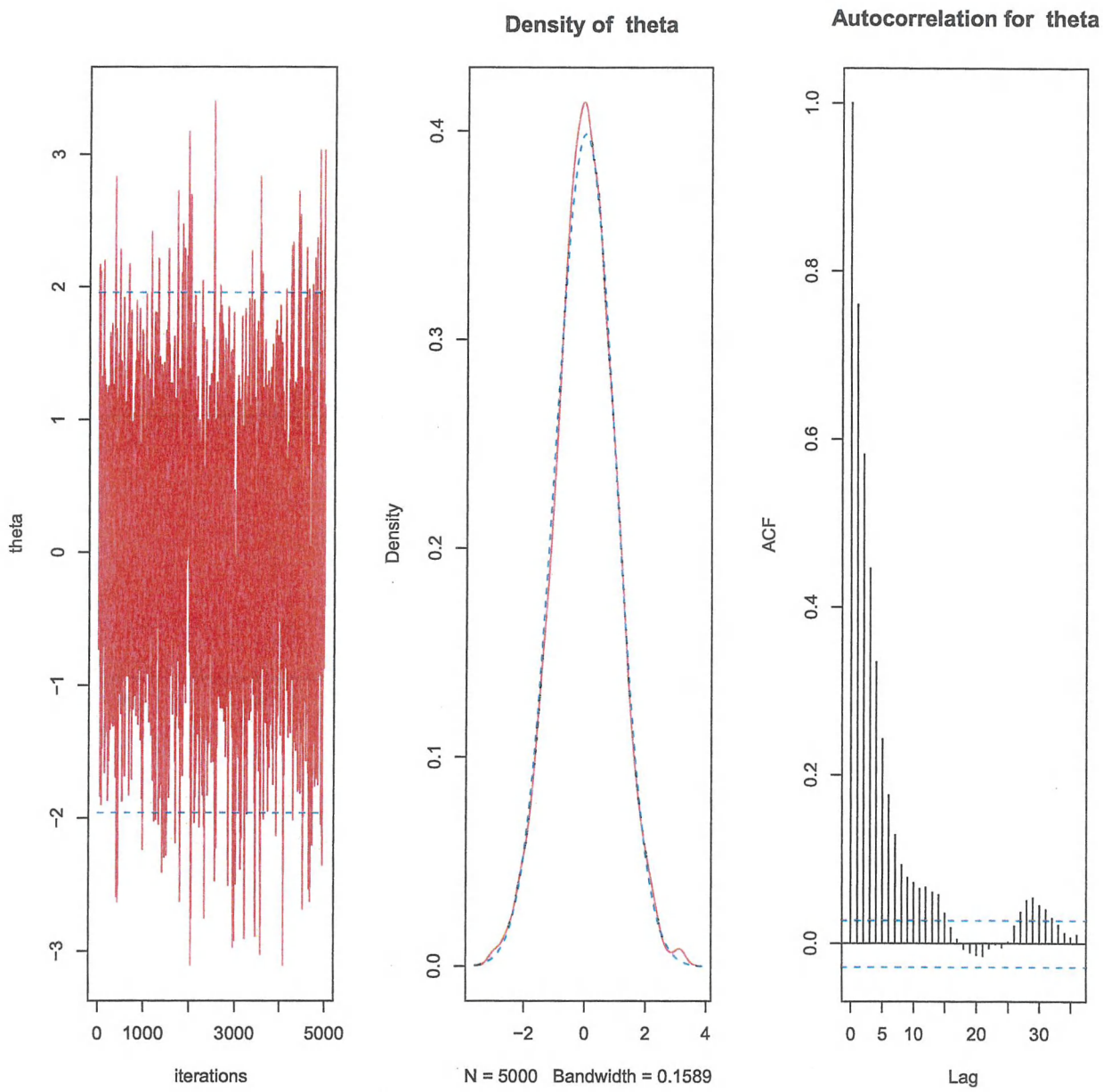

Figure 3: 5000 iterations from a Metropolis-Hastings sampler for $\theta \mid x \sim N(0,1)$ where the proposal distribution is normal with mean θ and chosen variance $\sigma^2 = 1$ and the starting point is $\theta^{(0)} = 1$. The first plot shows the trace plot of $\theta^{(t)}$, the second the observed density against the target density and the third the autocorrelation plot. The sample mean is 0.01340934 and the sample variance is 0.984014. 2.76% of points were observed to be greater than 1.96 and the acceptance rate is 0.7038.

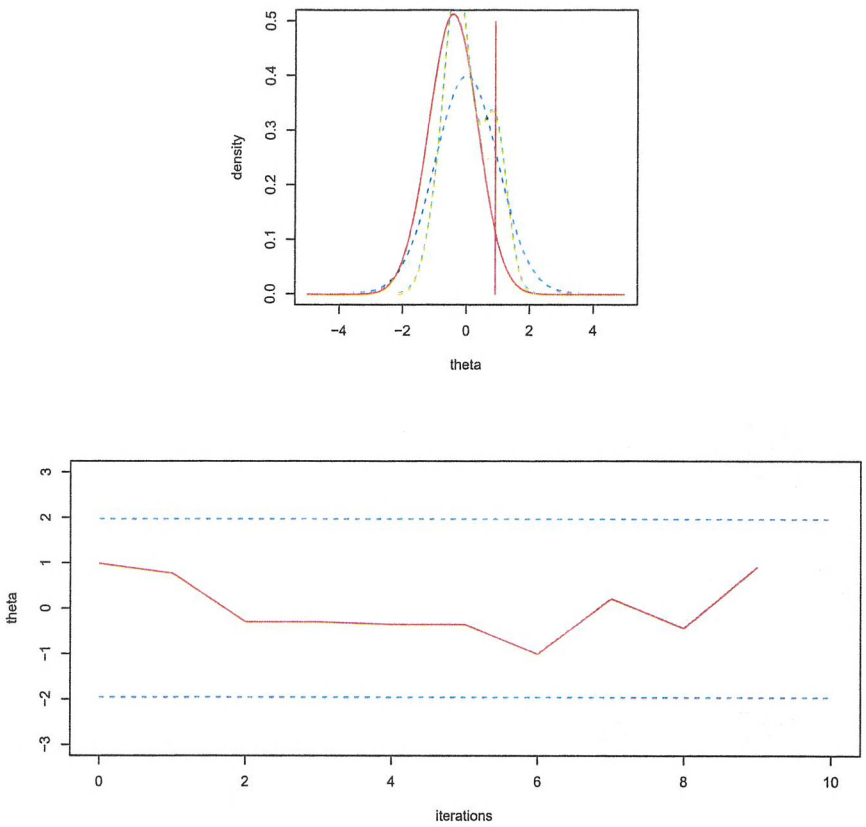

Figure 4: Nine iterations from a Metropolis-Hastings sampler for $\theta \mid x \sim N(0,1)$ where the proposal distribution is normal with mean θ and chosen variance $\sigma^2 = 0.36$ and the starting point is $\theta^{(0)} = 1$. The top plot shows the target distribution, $N(0,1)$, the proposal distribution for $\theta^{(9)}$ which is $N(-0.416, 0.36)$, the proposal $\theta^* = 0.923$ and the current observed density. The move is accepted; the bottom plot is the trace plot of $\theta^{(t)}$.

t	1	2	3	4	5	6	7	8	9
θ^*	0.786	−0.280	−1.092	−0.335	−1.169	−0.987	0.234	−0.416	0.923
Accept/Reject?	Accept	Accept	Reject	Accept	Reject	Accept	Accept	Accept	Accept
$\theta^{(t)}$	0.786	−0.280	−0.280	−0.335	−0.335	−0.987	0.234	−0.416	0.923

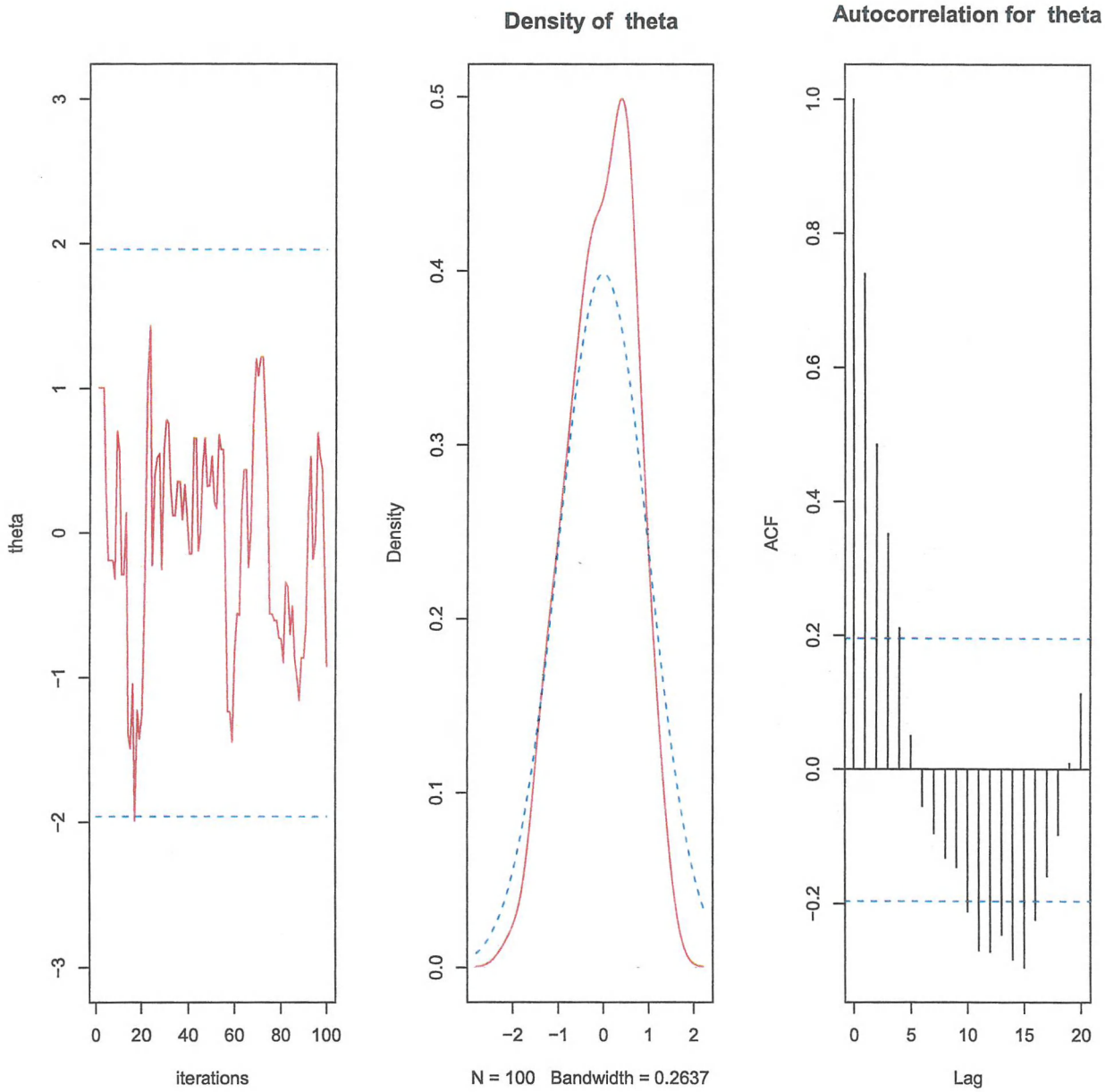

Figure 5: 100 iterations from a Metropolis-Hastings sampler for $\theta \,|\, x \sim N(0,1)$ where the proposal distribution is normal with mean θ and chosen variance $\sigma^2 = 0.36$ and the starting point is $\theta^{(0)} = 1$. The first plot shows the trace plot of $\theta^{(t)}$, the second the observed density against the target density and the third the autocorrelation plot. The sample mean is -0.04438661 and the sample variance is 0.5414882. 0% of points were observed to be greater than 1.96 and the acceptance rate is 0.83. Notice that, by comparing with Figure 2, reducing σ^2, that is we are proposing smaller jumps, increases the acceptance rate but reduces the mi ing.

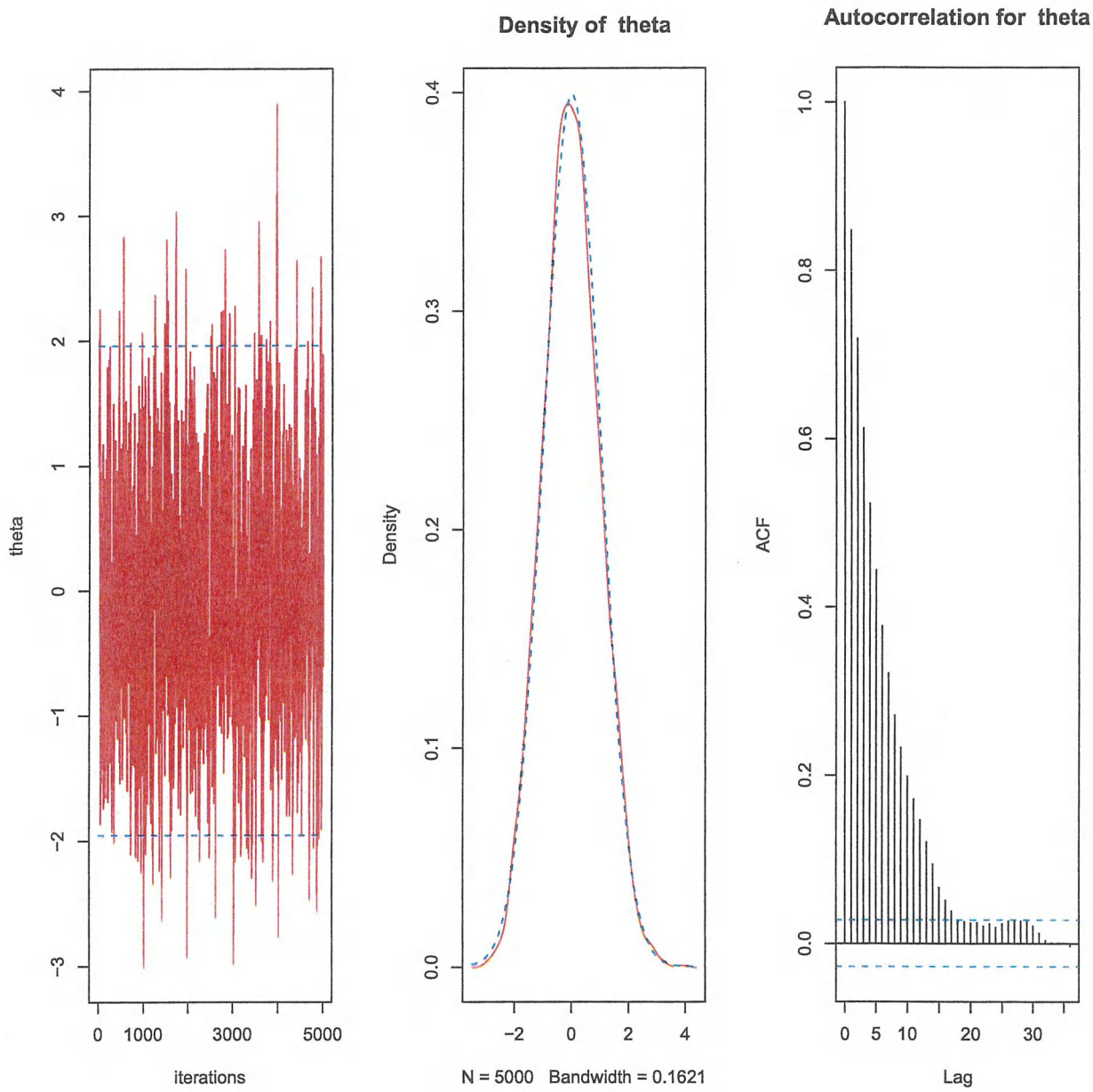

Figure 6: 5000 iterations from a Metropolis-Hastings sampler for $\theta \,|\, x \sim N(0,1)$ where the proposal distribution is normal with mean θ and chosen variance $\sigma^2 = 0.36$ and the starting point is $\theta^{(0)} = 1$. The first plot shows the trace plot of $\theta^{(t)}$, the second the observed density against the target density and the third the autocorrelation plot. The sample mean is -0.008158827 and the sample variance is 0.9792388. 2.42% of points were observed to be greater than 1.96 and the acceptance rate is 0.7894. Notice that, by comparing with Figure 3, reducing σ^2, that is we are proposing smaller jumps, increases the acceptance rate but reduces the mi ing.