

Part 3 homework: Markov Chain Monte Carlo

Marko Ivanovski

Github link: [here](#). All the relevant code can be found under homework5 directory.

Introduction

For the following task we were supposed to implement three types of sampling and then evaluate them on four different scenarios.

1. Scenario 1: bivariate standard normal with mean (0,0),
2. Scenario 2: the shape of the banana function with mean (0,0),
3. Scenario 3: the shape of the logistic regression likelihood using first two columns with mean (1.362, -0.707),
4. Scenario 4: the shape of the logistic regression likelihood using all columns with mean (2.048, -0.871, -0.498, 0.716, -0.083, -1.018, -0.819, 0.239, -0.605, -0.406, 0.477).

Rejection sampling

Rejection sampling was not hard to implement. The key parameter here is M , which I decided to compute since I noticed the algorithm was too slow if I plug some arbitrary big M . To calculate M I wrote a function that sampled from the proposal density 10 000 times and calculated the ratio between $\frac{f(x)}{g(x)}$ for all samples. Finally, M is the biggest number that is obtained and I added a small ϵ just to be safe that the assumption $Mg(x) \geq f(x)$ holds. I also implemented two proposal densities: multivariate normal and multivariate uniform distribution with constant height 1.

Scenario 1: bivariate standard normal

As non-tuned parameters, we can use the uniform proposal. The calculated M was 0.159. The results of the run are in Figure 1. The ACF was very low except for lag 0, the ESS was (1000,1000) and the mean of the samples was (0.013,-0.004). This is all expected since the samples are independent.

As tuned parameters, I used the standard bivariate normal distribution which is the same as the target distribution. Consequently, $M=1$ and the algorithm terminated faster. Again, all autocovariance lags were very low the ESS was (1000,1000) and the mean was (0.021,-0.010). Results are in Figure 2

Scenario 2: the shape of the banana function

As non-tuned parameters, the bivariate standard normal distribution was used. The calculated M was 31977.2. This is expected since the normal has a peak at (0,0) and the banana has a higher density elsewhere. Consequently, the algorithm ran very slow since it needed more time to propose points that

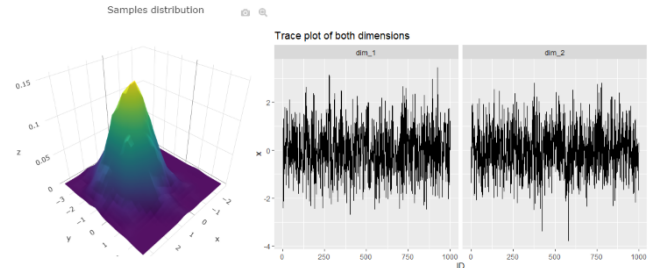


Figure 1. Sample distribution and trace plot from the samples obtained with rejection sampling on scenario 1 using multivariate uniform proposal.

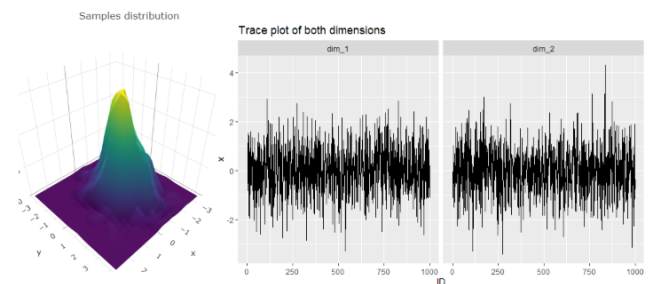


Figure 2. Sample distribution and trace plot from the samples obtained with rejection sampling on scenario 1 using bivariate standard normal proposal.

have a big density in the banana function. I only sampled 100 samples and it needed 10 minutes. The ESS was (100,100) as expected but the mean was (-0.062,3.327). The traceplot and the distribution are visualized in Figure 3. The problem with this proposal was that I had the variance of the normal proposal set to 0 and it did not propose enough points which are on the edges of the banana hence we only see the center of the banana.

Next, I tried sampling from the same distribution but this time using the uniform proposal. I expected better results here since the uniform proposal has a constant height which should converge faster than the normal proposal. The calculated M was 0.98. The results of the run are in Figure 4. The ESS of the samples was (1000,1000) and the mean was (0.12, 0.16) which is much better.

Scenario 3: the shape of the logistic regression likelihood using first two columns

The logistic regression density from which we tried sampling from at point $\beta = [\beta_0, \beta_1]$ was evaluated as:

$$L(\theta, x, y) = \prod_{i=1}^n \theta_i^{y_i} (1 - \theta_i)^{(1-y_i)}, \text{ where } \theta_i = \text{logit}^{-1}(\beta^T x_i)$$

I also took into account that β_0 is the intercept that corresponds to the first column of the dataset. This likelihood

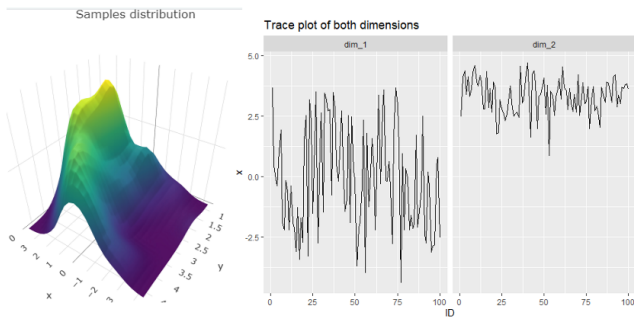


Figure 3. Sample distribution and trace plot from the samples obtained with rejection sampling on scenario 2 using standard bivariate normal proposal.

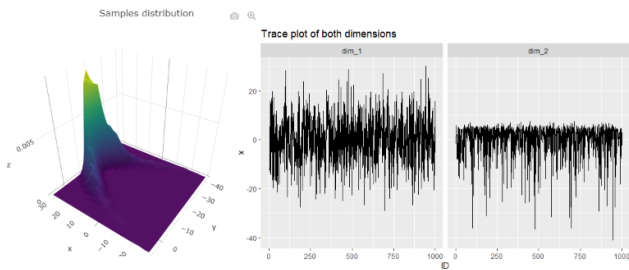


Figure 4. Sample distribution and trace plot from the samples obtained with rejection sampling on scenario 2 using multivariate uniform proposal.

returns very small numbers and when I used the bivariate standard normal distribution the optimal M turned out to be $1.147597e-212$. The algorithm wasn't fast which is why I ran it only on 100 samples. The ESS was (100,100) and the mean of the samples was (1.382, -0.714) which is close to the real mean. Results are in Figure 5.

I also tried to sample from this distribution using the uniform proposal but it was very slow so I decided not to include any results.

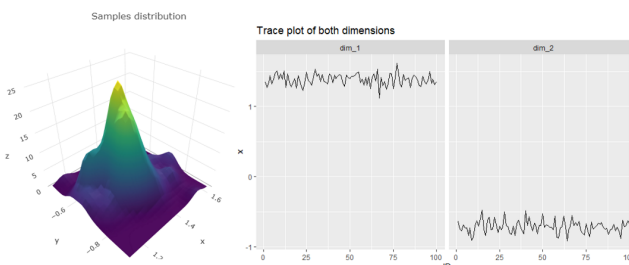


Figure 5. Sample distribution and trace plot from the samples obtained with rejection sampling on scenario 3 using standard bivariate normal proposal.

Metropolis-Hastings with Multivariate Normal proposal

Scenario 1: bivariate standard normal

As non-tuned parameters, I went with starting value at (0,0) and the covariance matrix was $0.5I_2$ (I_2 being the identity matrix with dimension 2). Results are in Figure 6. With these parameters, we can immediately see that the ACF plot shows big covariance for different lags which we don't want. The ESS was (12,23) for a sample of 1000 and the mean of the samples was (-0.52, -0.15). Traceplot chains also seem to be everywhere.

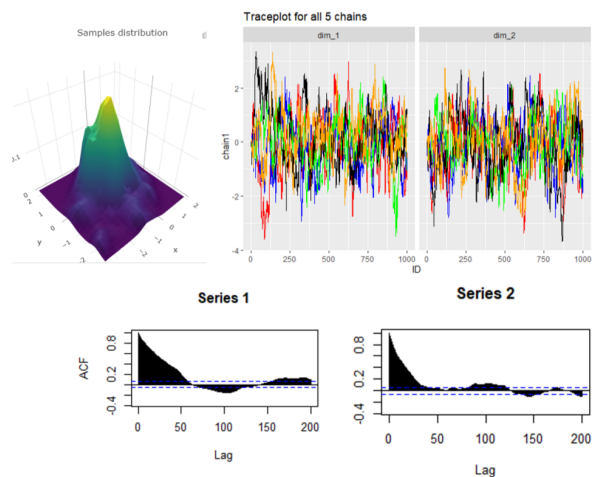


Figure 6. Sample distribution, trace plot and ACF plot for both dimension from the samples obtained with MH sampling on scenario 1 using bivariate normal proposal with covariance $0.5I_2$.

For the tuned parameters I chose starting position at (1,1) and covariance matrix $2I_2$. This should be better because we will propose states which are a bit further but still not by a lot. Results are in Figure 7. First, we see the ACF plot looks better since lags are lower. The ESS was (217, 145) and the mean of the samples was (-0.08, -0.04).

Scenario 2: the shape of the banana function

For the banana shape function, I could not manage to get good results. For the 'tuned' parameters the starting point was (5.5, 4) and the covariance matrix was $2I_2$. From Figure 8 we see the ACF plot does not look good. The sample ESS was (11,25) and the mean was (3.67,3.21). These results are poor for 1000 samples and if we tweak the parameters more we can get even worse results. Consecutive runs returned different results all the time (as seen from the traceplot).

Scenario 3: the shape of the logistic regression likelihood using the first two columns

Sampling from this distribution with MH was tricky since a lot of the mass is concentrated in a small area. For tuned parameters, I used the starting position (1.4, -0.76) and the covariance was $0.01I_2$. This gave good results (Figure 9) with low ACF and ESS: (166.85, 201.03). The mean of the samples was (1.36, -0.71), and running the chain multiple times produced similar results.

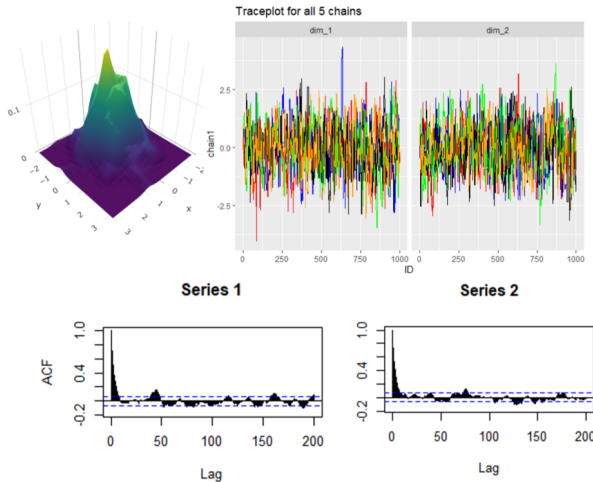


Figure 7. Sample distribution, trace plot and ACF plot for both dimension from the samples obtained with MH sampling on scenario 1 using bivariate normal proposal with covariance $0.5I_2$.

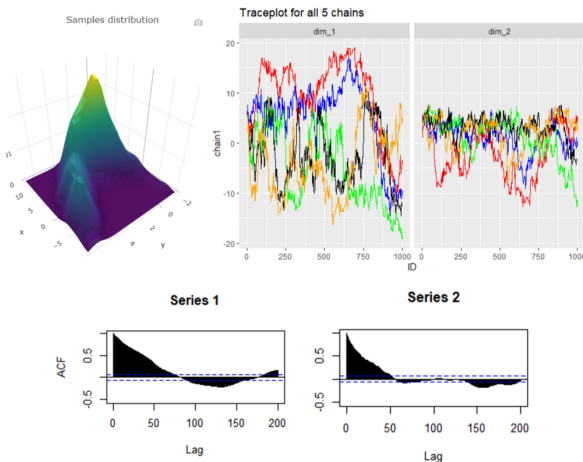


Figure 8. Sample distribution, trace plot and ACF plot for both dimension from the samples obtained with MH sampling on scenario 2 using bivariate normal proposal.

As non-tuned parameters, I choose a bigger covariance matrix $2I_2$. This results in ESS being smaller than 10 and the trace plot showed that the chain was stuck at the same point. This is expected since we are proposing points that are far and have a very low probability.

Scenario 4: the shape of the logistic regression likelihood using all columns

Sampling from this distribution was a challenge since the starting position played a huge role. If we are at some point with very low density it might be very hard to produce a point with a bigger density which is going to be far from our point since we are in 11 dimensions. I tried setting the initial point close to the real mean (2.05, -0.87, -0.5, 0.72, -0.09, -1.02, -0.82, 0.24, -0.61, -0.42, 0.47) and set a very small covariance $0.005I_{11}$. With these parameters I obtained ESS: (16, 143,

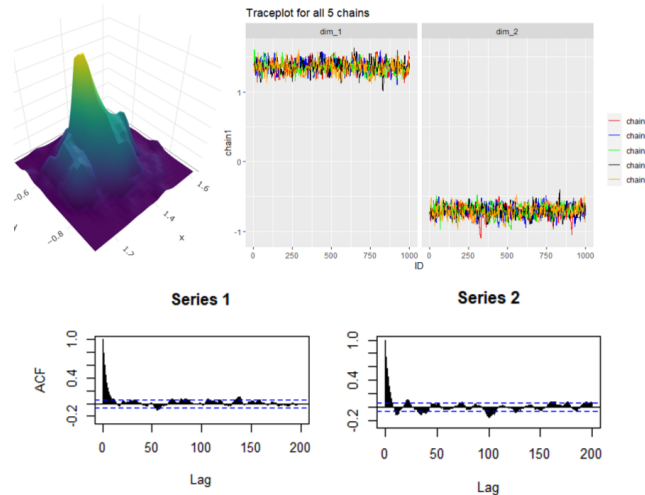


Figure 9. Sample distribution, trace plot and ACF plot for both dimension from the samples obtained with MH sampling on scenario 3 using bivariate normal proposal.

26, 26, 21, 18, 17, 10, 13, 22, 20) and mean: (2.04, -0.87, -0.49, 0.68, -0.06, -1.01, -0.81, 0.22, -0.60, -0.41, 0.46). To conclude, in this scenario I would consider doing a longer chain and trying burn-in to reduce the influence of the starting position.

Hamiltonian Monte Carlo

Scenario 1: bivariate standard normal

Sampling from the normal bivariate was not challenging, I used 10 leapfrog steps, $\epsilon=0.5$, and started at $q=(0.5,1)$. Results (Figure 10) suggest healthy samples, with ESS (415,713) and the mean (0.01,-0.02). For non-tuned parameters, we can just use any low L and some point far from our distribution to obtain high autocovariance and low ESS.

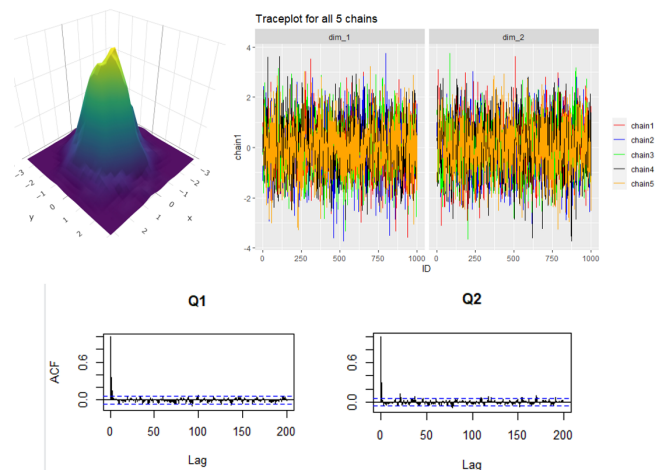


Figure 10. Sample distribution, trace plot and ACF plot for both dimension from the samples obtained with HMC sampling on scenario 1 with tuned parameters.

Scenario 2: the shape of the banana function

As tuned parameters for this scenario, I recycled the parameters we used during the lecture ($\epsilon = 0.6$, $L=27$, and starting $q=[0,0]$). The results (Figure 11) were good, rerunning the sampling generated 5 very similar chains and the ESS was a healthy (250, 295), the ACF plot looks good and the mean of the samples was (0.13, -0.13)

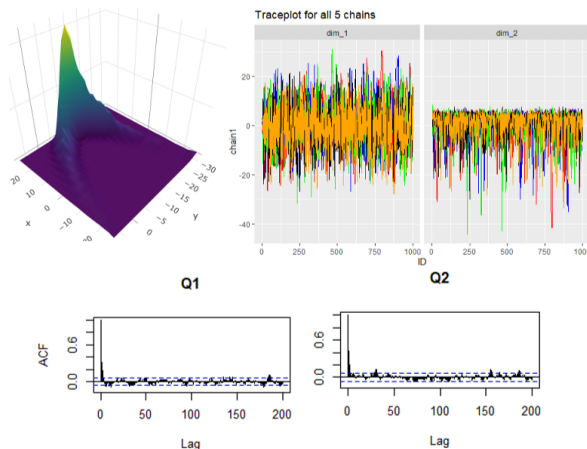


Figure 11. Sample distribution, trace plot and ACF plot for both dimension from the samples obtained with HMC sampling on scenario 2 with tuned parameters.

As non-tuned parameters, I reduced L to 4 so that the sampling would not have enough repetitions to generate the whole distribution. Figure 12 shows we did not generate half of the banana. Consequently, the ESS was low (4.5, 6.7) and the mean was (5.35, -0.73). The traceplot shows how all chains differ by a lot.

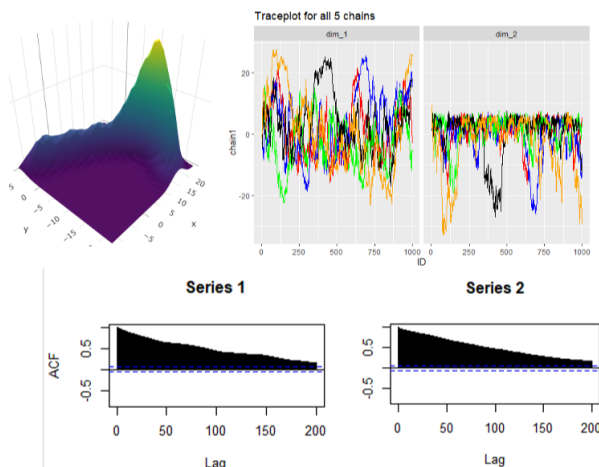


Figure 12. Sample distribution, trace plot and ACF plot for both dimension from the samples obtained with HMC sampling on scenario 2 with non-tuned parameters.

Scenario 3: the shape of the logistic regression likelihood using the first two columns

Sampling from this distribution was tricky because of tuning the right parameters. Generating the chain also took some

time. From Figure 13 we see that the traceplot shows that in some points we are staying for a long time which results in high autocovariance. Parameters used here were $L=60$, $\epsilon=0.001$, $q=(1.4, -0.58)$. The ESS was (19.2, 35.3) and the mean (1.39, -0.72).

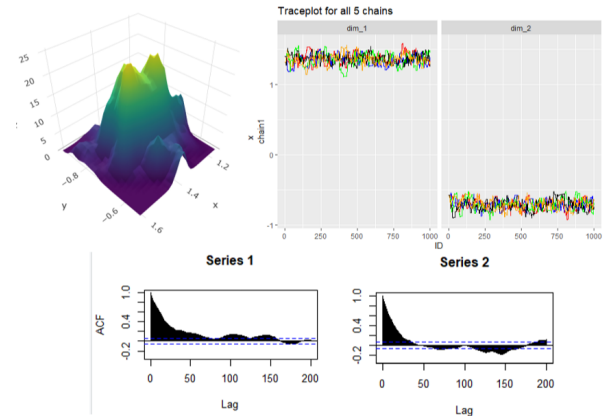


Figure 13. Sample distribution, trace plot and ACF plot for both dimension from the samples obtained with HMC sampling on scenario 3.

Scenario 4: the shape of the logistic regression likelihood using all columns

Lastly, for scenario 4 I started at $q=(2.05, -0.87, -0.5, 0.72, -0.09, -1.02, -0.82, 0.24, -0.61, -0.42, 0.47)$. Even if I moved a bit from this point the potential energy was 0. By using $L=50$ and $\epsilon=0.0009$ I got ESS: (10, 9, 10, 9, 9, 19, 7, 10, 11, 14, 10) and the mean: (2.21, -0.97, -0.52, 0.74, -0.015, -1.01, -0.86, 0.21, -0.51, -0.53, 0.50). The mean is not as close as what MH achieved.

Conclusion

We tried three different sampling techniques and saw their advantages and disadvantages. Rejections sampling seems to be ok for simple very low dimensional distributions for which we can find a great proposal. For scenario 1 we saw all methods worked well. HMC worked a bit better than MH but all were very close to the real mean. Scenario 2 demonstrated the flaws of the MH algorithm. In this case, HMC was by far superior to all since rejection sampling took a longer time. For the 3 scenario rejection sampling was super slow since the density is in a very small region. Between the other two, MH was better (closer to the mean, bigger ESS) at least in my case since I was able to tune the parameters better. Maybe with the correct tuning, we could get HMC to work better. Finally, in the final scenario 4, I wasn't able to find good parameters for both HMC and MH, although MH was a bit closer to the mean.