Homework 5: Markov Chain Monte Carlo

Matej Kalc

Introduction

Metropolis-Hastings, HMC, and rejection sampling are three methods for sampling from a proposal distribution. We will test those methods on four different scenarios: a bivariate standard normal, banana shaped function, logistic regression likelihood (using the first two columns) and logistic regression likelihood (using all columns). All code can be at https://github.com/KalcMatej99/M2HW5.

Metropolis-Hastings

The initial position and covariance matrix were calculated using the mean and variance of the samples from a trial run. For each scenario 5 chains were generated.

Bivariate standard normal

The mean of the samples returned from the tuned run is approximately (0,0), which is the true mean of this proposal (see Table 1). The fourth chain gave the best results since ess is the highest. When observing the movements of each dimension of each chain (see Figure 1), we see that all chains are centered at 0.

Chain	x	y	ess
1	-0.17	-0.02	(84.93 , 68.91)
2	-0.03	-0.09	(104.40, 98.07)
3	0.23	0.06	(114.04, 60.59)
4	0.06	0.01	(108.94, 133.58)
5	0.08	-0.02	(66.70, 84.15)

Table 1. Means of samples and ess for each chain generated using Metropolis-Hastings on the Bivariate proposal.

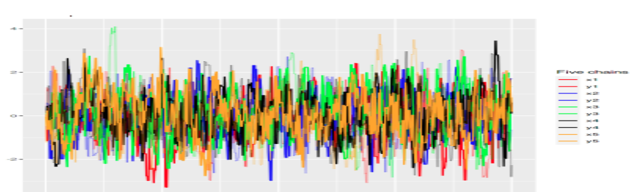

Figure 1. Trace plot of 5 chains generated using Metropolis-Hastings on the Bivariate proposal. Each color represents a chain. The first dimension is not dotted and the second is dotted.

Banana shape function

Metropolis-Hastings is not able to sample efficiently the banana shape function. The true mean is (0, 0). The returned sample means are not near the true mean (see Table 2). The ess for all five chains is low. The trace plot shows us how the chains moved in time (see Figure 2). We see that the chains are all different and they are not concentrated around a point (like in Figure 1).

Chain	x	y	ess
1	11.00	-4.31	(6.81, 7.10)
2	6.39	1.61	(11.11, 9.57)
3	-7.91	-1.91	(6.09, 5.90)
4	-2.22	-0.30	(6.98, 8.52)
5	-2.54	2.64	(11.27, 16.85)

Table 2. Means of samples and ess for each chain generated using Metropolis-Hastings on the banana shape function.

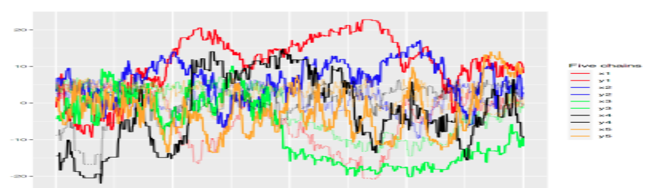

Figure 2. Trace plot of 5 chains generated using Metropolis-Hastings on the banana shape function. Each color represents a chain. The first dimension is not dotted and the second is dotted.

Logistic regression likelihood (using the first two columns)

The sample mean of all five chains is near the true mean (1.36, -0.70) (see Table 3). In the trace plot, we see that the first and second dimensions for all chains (x any y in Figure 6) are concentrated around different points.

Logistic regression likelihood (using all columns)

Metropolis-Hastings fails to generate efficient chains (see Table 4). The ess was not higher then 13. This method has difficulties with high dimension proposals. The reported sample means are not close to the ground truth (2.00, -0.84, -0.64, 0.72, -0.10, -0.85, -0.97, 0.23, -0.68, -0.42, 0.47).

Chain	x	y	ess
1	1.36	-0.70	(60.91, 51.81)
2	1.37	-0.70	(52.58, 94.49)
3	1.38	-0.70	(79.91, 76.36)
4	1.36	-0.68	(68.35, 51.07)
5	1.38	-0.70	(48.64 174.22)

Table 3. Means of samples and ess for each chain generated using Metropolis-Hastings on the logistic regression likelihood (using the first two columns).

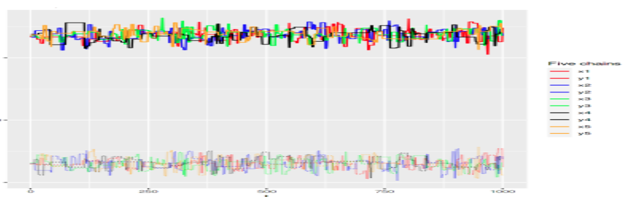

Figure 3. Trace plot of 5 chains generated using Metropolis-Hastings on the logistic regression likelihood (using the first two columns). Each color represents a chain. The first dimension is not dotted and the second is dotted.

Chain	1	2	3	4	5
x1	2.26	2.52	2.90	2.98	2.55
x2	-0.95	-0.88	-1.19	-1.11	-1.35
x3	-0.54	-0.76	-0.87	-0.92	-1.05
x4	0.70	0.68	1.63	1.15	1.18
x5	-0.06	-0.21	-0.15	-0.03	-0.13
x6	-1.29	-1.44	-1.91	-1.65	-0.68
x7	-0.41	-0.84	-1.16	-1.40	-1.16
x8	0.36	0.40	0.26	0.62	0.39
x9	-0.39	-0.39	-0.50	-1.29	-0.26
x10	-0.46	-0.53	-0.62	-0.85	-0.41
x11	0.55	0.62	0.73	0.50	0.69

Table 4. Means of samples for each chain generated using Metropolis-Hastings on the logistic regression likelihood (using all columns).

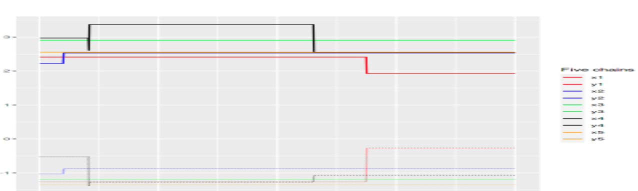

Figure 4. Trace plot of 5 chains generated using Metropolis-Hastings on the logistic regression likelihood (using all columns). Each color represents a chain. The first dimension is not dotted and the second is dotted. For clearer understating only the first two dimensions are reported.

HMC

The initial position was calculated using the mean of the samples from a trial run. For each scenario 5 chains were generated.

Bivariate standard normal

HMC gives better results since ess is higher than M-H and the sample mean is closer to the true mean (see Table 10). By inspecting the trace plot, we see that all chains are concentrated around 0 (see Figure 5).

Chain	x	y	ess
1	-0.03	-0.03	(1915.57, 1915.57)
2	0.01	0.01	(1993.15, 1993.15)
3	-0.03	-0.03	(2080.03, 2080.03)
4	0.02	0.02	(4854.22, 4854.22)
5	-0.01	-0.01	(3125.25, 3125.25)

Table 5. Means of samples and ess for each chain generated using HMC on the Bivariate proposal.

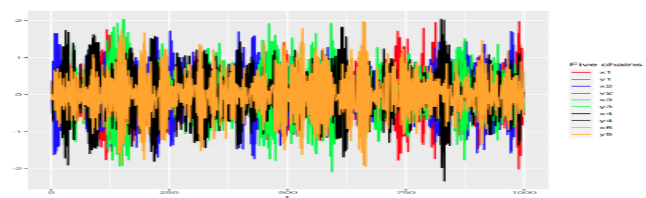

Figure 5. Trace plot of 5 chains generated using HMC on the Bivariate proposal. Each color represents a chain. The first dimension is not dotted and the second is dotted.

Banana shape function

The true mean of the banana shaped function is (0, 0). No chain is near the true mean (see Table 6), although the ess is much higher when compared whit M-H. Both methods have difficulties in sampling the tales of the banana. This could be solved by using an ideal starting position with HMC.

Chain	x	y	ess
1	-0.79	2.48	(213.78, 194.25)
2	-0.98	2.37	(299.81, 155.23)
3	-0.25	2.79	(386.61, 346.65)
4	-0.71	2.43	(244.88, 88.07)
5	-0.72	2.39	(321.81, 192.67)

Table 6. Means of samples and ess for each chain generated using HMC on the banana shape function.

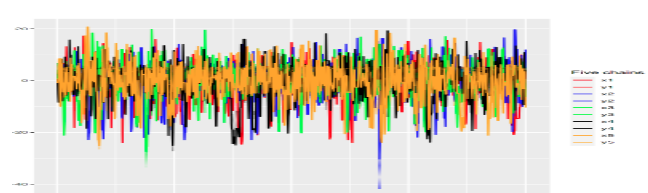

Figure 6. Trace plot of 5 chains generated using HMC on the banana shape function. Each color represents a chain. The first dimension is not dotted and the second is dotted.

Logistic regression likelihood (using the first two columns)

Using HMC improved the ess and the sample means are closer to the true means (see Table 7).

Chain	x	y	ess
1	1.36	-0.71	(1454.17, 1473.93)
2	1.36	-0.71	(1146.17, 1134.92)
3	1.36	-0.71	(1323.11, 1342.86)
4	1.36	-0.71	(2472.01, 2523.44)
5	1.36	-0.71	(1106.41, 1113.44)

Table 7. Means of samples and ess for each chain generated using HMC on the logistic regression likelihood (using the first two columns).

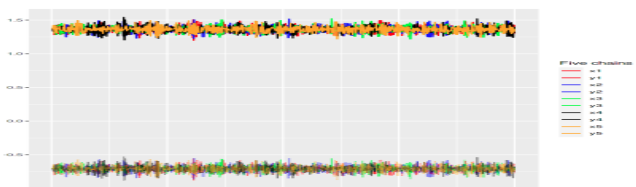

Figure 7. Trace plot of 5 chains generated using HMC on the logistic regression likelihood (using the first two columns). Each color represents a chain. The first dimension is not dotted and the second is dotted.

Logistic regression likelihood (using all columns)

Again HMC increased the ess and the sample means are closer to the ground truth. The ess of chain 3 is (470.98, 417.99, 595.14, 414.13, 427.03, 381.21, 247.24, 471.26, 331.60, 220.16, 510.34).

Chain	1	2	3	4	5
x1	2.05	2.05	2.04	2.04	2.05
x2	-0.87	-0.87	-0.87	-0.87	-0.87
x3	-0.50	-0.50	-0.50	-0.50	-0.50
x4	0.71	0.71	0.71	0.72	0.71
x5	-0.08	-0.08	-0.08	-0.08	-0.08
x6	-1.01	-1.01	-1.01	-1.02	-1.02
x7	-0.82	-0.82	-0.81	-0.82	-0.82
x8	0.23	0.23	0.24	0.24	0.24
x9	-0.60	-0.60	-0.60	-0.60	-0.60
x10	-0.40	-0.40	-0.40	-0.40	-0.40
x11	0.47	0.47	0.47	0.48	0.47

Table 8. Means of samples for each chain generated using HMC on the logistic regression likelihood (using all columns).

Rejection sampling

The selected 2D envelope is a multivariate standard normal distribution. M is the most important feature. M was calculated using a trial run with 1000 samples. For each sample

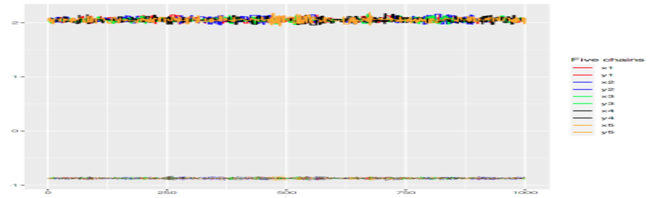

Figure 8. Trace plot of 5 chains generated using HMC on the logistic regression likelihood (using all columns). Each color represents a chain. The first dimension is not dotted and the second is dotted. For clearer understating only the the first two dimensions are reported.

$\frac{f(x)}{g(x)}$ was calculated and the maximum of those ratios was taken as M. In Table 9 we see that calculated M's for each chain and proposal.

	M			
Chain	Bivariate	Banana	Log. (2 col.)	Log.
1	1	169.24	1.89e-213	1e-300
2	1	5.65	8.74e-214	1.29e-255
3	1	96.93	1.90e-216	1e-300
4	1	190.03	2.93e-215	1.96e-276
5	1	22.60	5.01e-215	1e-300

Table 9. M for each chain and proposal.

Bivariate standard normal

Using rejection sampling we achieve good, high ess and the sample means are close to the real mean. During execution we notice that rejection sampling is much slower in comparison with the previous.

Chain	x	y	ess
1	0.01	-0.07	(1000, 1000)
2	0.05	0.01	(1000, 1000)
3	-0.01	0.01	(1000, 1000)
4	0.02	0.01	(1000, 1068.274)
5	-0.01	-0.04	(1000, 1000)

Table 10. Means of samples and ess for each chain generated using HMC on the Bivariate proposal.

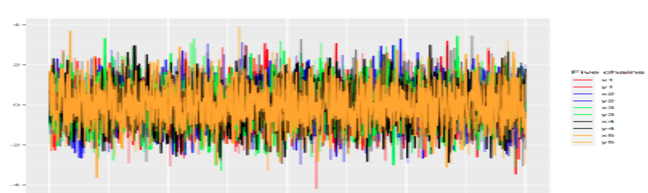

Figure 9. Trace plot of 5 chains generated using rejection sampling on the Bivariate proposal. Each color represents a chain. The first dimension is not dotted and the second is dotted.

Banana shape function

Rejection sampling could not produce a chain with a sample mean close to the true mean (see Table 11), even though it has high ess.

Chain	x	y	ess
1	0.02	3.21	(1000, 1000)
2	0.01	2.80	(1000, 1000)
3	0.06	2.83	(1000, 1000)
4	0.11	3.36	(1000, 1000)
5	0.00	2.72	(1000, 1000)

Table 11. Means of samples and ess for each chain generated using rejection sampling on the banana shape function.

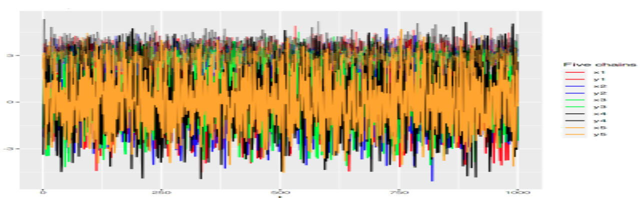

Figure 10. Trace plot of 5 chains generated using rejection sampling on the banana shape function. Each color represents a chain. The first dimension is not dotted and the second is dotted.

Logistic regression likelihood (using the first two columns)

Rejection sampling generated chains with high ess as with HMC. Again we consider that rejection sampling is much slower than HMC. For this particular scenario, HMC would be the clear method to use.

Chain	x	y	ess
1	1.37	-0.71	(1000, 893.13)
2	1.37	-0.71	(1000, 1000)
3	1.37	-0.71	(1000, 1000)
4	1.36	-0.71	(1000, 1000)
5	1.36	-0.71	(1000, 1000)

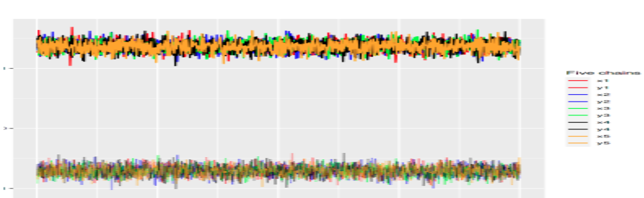

Figure 11. Trace plot of 5 chains generated using rejection sampling on the logistic regression likelihood (using the first two columns). Each color represents a chain. The first dimension is not dotted and the second is dotted.

Logistic regression likelihood (using all columns)

Since rejection sampling is slower than M-H and HMC, we used only 100 samples. Comparing results from Table 12 with HMC, the results are worse. Also if we inspect the trace plot (see Figure 12) we see that each chain is not always consistent as in Figure 8.

Chain	1	2	3	4	5
x1	1.38	1.42	1.41	1.49	1.51
x2	-0.43	-0.56	-0.55	-0.54	-0.61
x3	-0.31	-0.25	-0.34	-0.34	-0.29
x4	0.50	0.35	0.45	0.42	0.39
x5	-0.07	-0.04	-0.08	-0.06	-0.07
x6	-0.75	-0.68	-0.65	-0.70	-0.60
x7	-0.45	-0.47	-0.55	-0.43	-0.47
x8	0.27	0.11	0.05	0.12	0.15
x9	-0.32	-0.42	-0.31	-0.45	-0.24
x10	-0.23	-0.30	-0.34	-0.18	-0.16
x11	0.32	0.29	0.28	0.38	0.29

Table 12. Means of samples for each chain generated using rejection sampling on the logistic regression likelihood (using all columns).

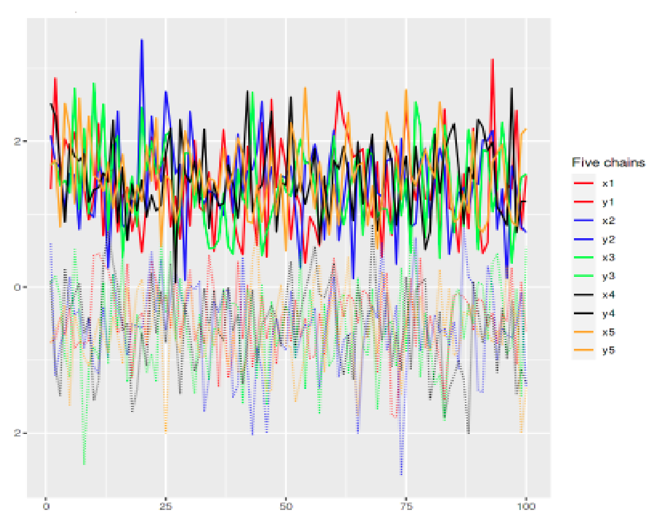

Figure 12. Trace plot of 5 chains generated using rejection sampling on the logistic regression likelihood (using all columns). Each color represents a chain. The first dimension is not dotted and the second is dotted. For clearer understating only the the first two dimensions are reported.

Discussion

Metropolis-Hastings performed well with the bivariate distribution and logistic regression likelihood (with the first two columns). In the other two scenarios, the method failed. HMC performed better than M-H in each scenario. Rejection sampling is slower than the other two methods. Of the three HMC is the best method, if we consider the performance and execution time. The majority of parameters were set using a trial run. Other parameters, such as epsilon in HMC were using trial and error. The parameters weren't hard to tune.