

# Homework MCMC

ANDRAŽ DE LUISA

A)  $X \sim N(0, 1)$ ,  $f(x) = 10 \exp(-5(x-3)^4)$

1.  $I = E_X[f(x)] = \int_{-\infty}^{\infty} f(x)p(x)dx = 0.089$ , with absolute error  $< 2e-5$ , where  $p(x)$  is the standard normal distribution density.
2. Monte-Carlo approximation (100 samples):  $I = 0.062$  with  $SE = 0.053$ .
3. Repeat the Monte-Carlo approximation 1000 times: for each iteration we construct a 90% confidence interval (using the normal approximation) and check if the true value is contained in it. It is contained in the 66% of cases. We get such a low percentage because the normal assumption doesn't hold (function values are mostly 0 with some high peaks). However, the final approximation (the average of the 1000 iterations) is quite precise:  $I = 0.091$  with  $SE = 0.002$ .
4. Importance sampling: we want the surrogate distribution to have a similar shape to  $f(x) * p(x)$  and to be easy to sample from

$$f(x) * p(x) = \frac{10}{\sqrt{2\pi}} \exp(-(5(x-3)^4 + x^2/2)) \propto \exp(-5(x-3)^4 + x^2/2),$$

therefore we choose the density  $g(x)$  of the normal distribution  $N(2.5, 0.25)$ . The condition  $f(x) * p(x) \gg 0 \implies g(x) \gg 0$  is clearly satisfied (the 4-th order polynomial in  $f(x)$  decreases its values towards 0 much faster than the 2-nd order one in  $g(x)$ ), hence we expect the variance to be lower than the simple MC estimator. This is actually the case, since the estimated variance in (2) is 0.285, while the variance of the importance sampling estimator is 0.002. The obtained estimation is  $I = 0.098$ .

5. Rejection sampling with (standard) logistic distribution based envelope:  $env(x) = \frac{e^{-x}}{(1+e^{-x})^2}$ . The appropriate scaling parameter  $M$  is computed through the inequality  $M > p(x)/env(x) \forall x \implies M > 4/\sqrt{2\pi}$ . The integral approximation is  $I = 0.034$  with its variance being 0.115. It is smaller than that in (2), but this should hold just by chance (we only change the sampling procedure, which is probably not as efficient as the default one).

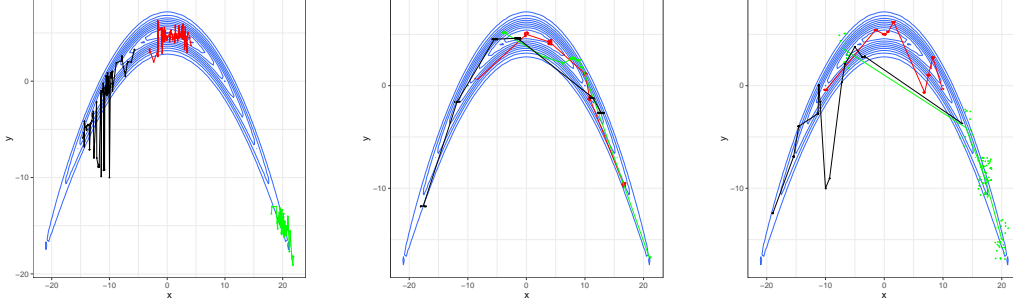


Figure 1: Contour plot of the distribution density with chains starting from 3 different starting points  $x_1 = (0, 5)$ ,  $x_2 = (20, -15)$  and  $x_3 = (-10, -10)$ .

6. Metropolis-Hastings with a  $U(x_i - \delta, x_i + \delta)$ ,  $\delta > 0$  distribution. The rejection rate monotonically increases with  $\delta$  increasing. The variance of the estimator tends toward 0 for  $\delta$  approaching 0 and for  $\delta \gg 0$ , with a peak around  $\delta \in [3, 7]$ . For  $\delta = 5$  the estimation is  $I = 0.168$  with variance 0.985 (much bigger than in (2)). The effective sample size is only 29 (this is the main cause of the increased variance).

B) Handwritten

C) Sample from a distribution, whose density is proportional to  $f_n$ , using Metropolis-Hastings algorithm. In Figure 1 the first 100 samples of the obtained chains are shown over the true distribution density. In Table 1 the rejection rates and effective sample sizes are presented. The proposal distribution 3 is a multivariate normal distribution, with the mean being the previous point, and a diagonal covariance matrix. It is not really better than the other two. The shape of the desired distribution is causing the extremely low efficiency, since the correlations between the coordinates are changing with respect to the position of the point.

D) Metropolis-Hastings sampler for Bayesian logistic regression

- The function, proportional to the posterior distribution is the product of the normal density (with  $\sigma^2 = 100$ ) and the logistic regression likelihood (with an added intercept parameter).
- In Table 2 the rejection rates and effective sample sizes for each of the 4 independent chains are presented. Some concerns arise because

1	Rej. rate	ESS (x)	ESS (y)	2	Rej. rate	ESS (x)	ESS (y)
$x_1$	0.171	13.8	46.2	$x_1$	0.937	13.3	16.3
$x_2$	0.344	3.1	3.0	$x_2$	0.952	16.4	20.2
$x_3$	0.298	1.8	1.6	$x_3$	0.938	5.1	12.4

3	Rej. rate	ESS (x)	ESS (y)
$x_1$	0.925	17.5	32.2
$x_2$	0.927	21.4	64.5
$x_3$	0.937	5.48	6.0

Table 1: Rejection rates and effective sample sizes for the tree proposal distributions.

	Rej. rate	ESS ( $\beta_1$ )	ESS ( $\beta_2$ )	ESS ( $\beta_3$ )	ESS ( $\beta_4$ )	ESS ( $\beta_5$ )	ESS ( $\beta_6$ )
$x_1$	0.979	36.0	74.1	72.8	41.2	113.0	102.2
$x_2$	0.984	29.7	53.9	76.7	49.2	63.6	114.0
$x_3$	0.981	21.4	43.5	60.4	73.9	38.4	55.2
$x_4$	0.977	53.4	45.6	110.3	89.0	79.4	58.9
best	0.867	196.8	160.4	475.7	134.7	207.7	104.2

Table 2: Starting points are independently normally distributed around 0. The irrelevant parameters (by construction of the toy dataset) are  $\beta_4$  and  $\beta_5$ , while  $\beta_6$  is the intercept coefficient. The best-performing sampler is included for easier comparisons.

of the extremely high rejection rates (with high autocorrelations and small ESS direct consequences). Note for example that on 10000 samples, we get ESSes as small as 20.

- The covariance matrix of the best performing sampler is  $\text{diag}(.5, .5, .5, .1, .1, .1)$ . The results are presented in the last line of Table 2. The efficiency is improved, but it still isn't great. In Figure 2 a comparison between traceplots is shown. We can clearly see how often is the bad sampler stuck in some point (rejecting the new proposals).
- In Table 3 the estimates of the parameters, obtained from the posterior distribution, are shown. Additionally, the posterior probability  $P(|\beta_3| > \frac{1}{10} | data) = 0.80$  with  $SE = 0.003$ .

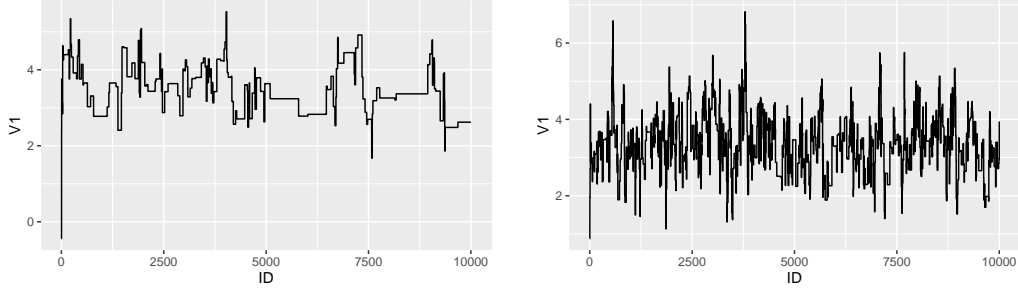


Figure 2: Traceplots for the parameter  $\beta_1$  from the 2nd chain of the bad sampler (left) and from the best sampler (right).

	estimate	standard error	ground truth
$\beta_1$	3.41	0.05	3
$\beta_2$	2.23	0.04	-2
$\beta_3$	-0.19	0.01	0.1
$\beta_4$	-0.00	0.03	0
$\beta_5$	0.49	0.02	0
$\beta_6$	-0.69	0.03	0

Table 3: The obtained estimates are not accurate (notice for example the estimate for  $\beta_2$ , which is completely wrong). I suspect the irrelevant parameters have a really bad influence on the results.