Project 1: Markov chain Monte Carlo techniques

Deadline: Friday, February 6th. Please send your solutions to andrea.riebler@ntnu.no.

General comments:

- This project is compulsory. To be able to take the exam, a reasonable attempt must be made to solve all problems.

- The project report should be formulated as a scientific report. In particular, it should be possible to understand what you have done without reading the questions in this problem text. Moreover, the text in the project report should consists of full sentences and proper punctuation should by used throughout, also in equations! All results you present should be discussed. What can you (and the world) learn from your results? The project text should be written so that it is easy to follow by your fellow students in MA8702. The project report should consist of one (and only one) pdf-file. It should include derivation of formulas that you are using in your implementations. The project report should also include the code you have used to solve the project and the plots you have generated. Associated to the various plots there should be captions explaining the contents of the plots, and in addition all the plots should be referenced, explained and discussed in the main text of the report

  The report can be written in English or Norwegian. The project can be done alone or in groups of two or three persons. If you do the project in a group, only one of the individuals in the group should hand in the solution via email to andrea.riebler@ntnu.no. In your solution, specify your (full) names, NOT student or candidate numbers.

1 Metropolis-Hastings (MH) for bivariate densities

We will consider three different bivariate target densities for $\boldsymbol{x} = (x_1, x_2)^t$:

1. Standard Gaussian distribution with correlation:
$$\pi(\boldsymbol{x}) = \frac{1}{2\pi|\boldsymbol{\Sigma}|^{1/2}} \exp\left(-\frac{1}{2}\boldsymbol{x}^t\boldsymbol{\Sigma}^{-1}\boldsymbol{x}\right),$$
where $\boldsymbol{\Sigma}$ has 1 on the diagonal and $\rho = 0.9$ on the off-diagonal.

2. A multimodal density:
$$\pi(\boldsymbol{x}) = \sum_{i=1}^{3} w_i \frac{1}{2\pi|\boldsymbol{\Sigma}_i|^{1/2}} \exp\left(-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu}_i)^t\boldsymbol{\Sigma}_i^{-1}(\boldsymbol{x}-\boldsymbol{\mu}_i)\right),$$
which is a mixture of Gaussian densities where weights are $w_1 = 1/3$, $w_2 = 1/3$ and $w_3 = 1/3$, means are $\boldsymbol{\mu}_1 = (-1.5, -1.5)^t$, $\boldsymbol{\mu}_2 = (1.5, 1.5)^t$ and $\boldsymbol{\mu}_3 = (-2, 2)^t$, and covariance matrices $\boldsymbol{\Sigma}_i$ all have correlation 0 and variance $\sigma_1^2 = 1$, $\sigma_2^2 = 1$ and $\sigma_3^2 = 0.8$.

3. A volcano-shaped density:
$$\pi(\boldsymbol{x}) \propto \frac{1}{2\pi} \exp\left(-\frac{1}{2}\boldsymbol{x}^t\boldsymbol{x}\right)(\boldsymbol{x}^t\boldsymbol{x} + 0.25),$$

We will explore each one with random walk Metropolis–Hastings (MH) algorithms, Langevin MH and Hamiltonian MH.

1.1 Plotting

Visualize the three densities on a grid covering $[-5, 5] \times [-5, 5]$. The grid spacing could be 0.1, which gives $101 \times 101 =$ grid cells. Note that the volcano-density in 3. is not normalized, but the relative levels are still representative.

1.2 Random walk MH

- Implement a random walk MH sampler for the Gaussian density in 1. above. Try tuning parameter $\sigma = 0.5$ in the random walk proposal, and then experiment with a few others. Keep track of the mean acceptance rate while the algorithm is running. Inspect the autocorrelation of the resulting Markov chain, and use this along with trace plots of x_1 and x_2 to evaluate the suitability of the different levels of the tuning parameter. Please discuss which value of σ you would choose to get close to the theoretically preferred acceptance rate?
- Implement a random walk MH sampler for the multimodal density in 2. above. Repeat the same algorithm evaluation as for the Gaussian density above.
- Implement a random walk MH sampler for the volcano-shaped density in 3. above. Repeat the same algorithm evaluation as above.

1.3 Langevin MH

- Implement a Langevin MH sampler for the Gaussian density in 1. above. Try tuning parameter $\sigma = 0.5$ in the Langevin proposal, and then experiment with a few others. Keep track of the mean acceptance rate while the algorithm is running. Inspect the autocorrelation of the resulting Markov chain, and use this along with trace plots of x_1 and x_2 to evaluate the suitability of the different levels of the tuning parameter. Compare also with the Random Walk MH in the previous section. Keep in mind that each iteration of the Langevin MH sampler relies on the evaluations of both target and derivative. Please discuss which value of σ you would choose to get close to the theoretically preferred accptance rate?

- Implement a Langevin MH sampler for the multimodal density in 2. above. Repeat the same algorithm evaluation as for the Gaussian density above.

- Implement a Langevin MH sampler for the volcano-shaped density in 3. above. Repeat the same algorithm evaluation as above.

1.4 Hamiltonian MH

- Implement a Hamiltonian MH sampler for the Gaussian density in 1. above. Set the momentum proposal to $\boldsymbol{z} \sim N(0, \boldsymbol{I})$. Try tuning parameter $\epsilon = 0.1$ in the leap-frog scheme for $T = 10$ steps, and then experiment with a few other settings. Keep track of the mean acceptance rate while the algorithm is running. Approximate the integrated autocorrelation of the resulting Markov chain, and use this along with trace plots of x_1 and x_2 to evaluate the suitability of the different levels of the tuning parameter. Compare also with the Random Walk and Langevin MH in the previous sections. Keep in mind that each iteration of the Hamiltonian MH sampler requires several evaluations of both target and derivative in the leap-frog calculation.

- Implement a Hamiltonian MH sampler for the multimodal density in 2. above. Repeat the same algorithm evaluation as for the Gaussian density above.

- Implement a Hamiltonian MH sampler for the volcano-shaped density in 3. above. Repeat the same algorithm evaluation as above.

2 RStan: Gamma-Poisson Hierarchical model

We consider an example from George, et al. (Scandinavian Journal of Statistics, 20:147–156, 1993) concerning the number of failures of ten power plants. The data are as follows:

Pump	1	2	3	4	5	6	7	8	9	10
y	5	1	5	14	3	19	1	1	4	22
t	94.3	15.7	62.9	126.0	5.24	31.4	1.05	1.05	2.1	10.5

Here, y_i is the number of times that pump i failed and t_i is the operation time of the pump (in 1000s of hourse). Pump failures are modelled as:

$$y_i \mid \lambda_i \sim \text{Poisson}(\lambda_i t_i)$$

Conjugate prior for λ_i:

$$\lambda_i \mid \alpha, \beta \sim \text{Gamma}(\alpha, \beta)$$

Hyper-prior on α and β:

$$\alpha \sim \text{Exp}(1.0) \qquad\qquad \beta \sim \text{Gamma}(0.1, 1.0)$$

- Install RStan on your computer. For instructions see `https://github.com/stan-dev/rstan/wiki/RStan-Getting-Started`
- Write a Stan program for the model presented above and save it in "pump.stan". Be sure that your Stan programs ends in a blank line without any characters including spaces and comments.
- Prepare the data in R and fit the model in R using the function `stan` and save the results in an object called `fit`.
- Check the effective sample size (ESS) estimates and traceplots to assess convergence. Inspect and interpret your posterior output.