Assignment 1
Miodrag Bolic

Due: February 8, 2021
Total number of points is 10. Please solve 3 out of 4 problems.

Instructions:
Upload your answers in a ipynb notebook to UOttawa Bright Space.

Your individual submissions should use the following filenames:
ELG_5218_YOURNAME_HW1.ipynb

Your code should be in code cells as part of your notebook. Do not use any different format.

*Do not just send your code. The homework solutions should be in a report style. Be sure to add comments to your code as well as markdown cells where you describe your approach and discuss your results. *

Please submit your notebook in an executed status, so that we can see all the results you computed. However, we will still run your code and all cells should reproduce the output when executed.

If you have multiple files (e.g. you've added code files or images) create a tarball for all files in a single file and name it: ELG_5218_YOURNAME_HW1.tar.gz or ELG_5218_YOURNAME_HW1.zip

Problem 1: Modeling

This problem is from the book: Bayesian Cognitive Modeling: A Practical Course by M. D. Lee and E.-J. Wagenmakers. It is about the Bayesian approach in accessing autocorrelation coefficient.

Read sections 5.1 and 5.2 from this link: https://cpb-us-e2.wpmucdn.com/faculty.sites.uci.edu/dist/0/180/files/2011/03/BB_Free.pdf

Simulate models explained in Sections 5.1 and 5.2. Please note, that Gaussian distribution N(0,001) means mean of 0 and precision of 0.001. Instead of sampling from Inverse Gamma in order to obtain sigma, you should sample from Gamma distribution and obtain precisions. Please make sure that the parameters you select for Gamma distribution are uninformative. The observed data take the form $\boldsymbol{x_i} = (x_{i1}, x_{i2})$.

A) Section 5.1: Two data sets are given, where each pair of elements is separated by ;.

Dataset 1: x = {0.8,102; 1,98; 0.5,100; 0.9,105; 0.7,103; 0.4,110; 1.2,99; 1.4,87; 0.6,113; 1.1,89; 1.3,93}
Dataset 2: x = {0.8,102; 1,98; 0.5,100; 0.9,105; 0.7,103;0.4,110; 1.2,99; 1.4,87; 0.6,113; 1.1,89; 1.3,93; 0.8,102; 1,98; 0.5,100; 0.9,105; 0.7,103; 0.4,110; 1.2,99; 1.4,87; 0.6,113; 1.1,89; 1.3,93}

i) Simulate the model in Figure 5.1 using Julia Turing (other acceptable options are Gen in Julia or Pyro or NumPyro in Python).

ii) Reproduce Figure 5.2 and do Exercise 5.1.1 and 5.1.2

B) Section 5.2:

i) Simulate the model in Figure 5.3 using Julia Turing (other acceptable options are Gen in Julia or Pyro or NumPyro in Python).

ii) Reproduce Figure 5.4 and do Exercises 5.2.1 - 5.2.4.

Problem 2: Regression

This problem is from Bayesian Learning by Mattias Villani
https://github.com/mattiasvillani/BayesLearnCourse

Please download data set tempLinkoping.txt from
https://github.com/mattiasvillani/BayesLearnCourse/blob/master/Labs/TempLinkoping.txt .
Please write program in Julia Turing (other acceptable options are Gen in Julia or Pyro or NumPyro in Python) to solve this problem.

The dataset `tempLinkoping` contains daily temperatures (in Celcius degrees) at Malmslätt, Linköping over the course of the year 2016 (366 days since 2016 was a leap year). The response variable is *temp* and the covariate is

$$time = \frac{\text{the number of days since beginning of year}}{366}.$$

The task is to perform a Bayesian analysis of a quadratic regression

$$temp = \beta_0 + \beta_1 \cdot time + \beta_2 \cdot time^2 + \varepsilon,\ \varepsilon \overset{iid}{\sim} N(0, \sigma^2).$$

a)

Write a program that *simulates from the joint posterior distribution* of β_0, β_1,β_2 and σ^2. Plot the marginal posteriors for each parameter as a histogram. Also produce another figure with a scatter plot of the temperature data and overlay a curve for the posterior median of the regression function $f(time) = \beta_0 + \beta_1 \cdot time + \beta_2 \cdot time^2$, computed for every value of *time*. Also overlay curves for the lower 2.5% and upper 97.5% posterior credible interval for $f(time)$. That is, compute the 95% equal tail posterior probability intervals for every value of *time* and then connect the lower and upper limits of the interval by curves. Does the interval bands contain most of the data points? Should they?

b)

It is of interest to locate the *time* with the highest expected temperature (that is, the *time* where $f(time)$ is maximal). Let's call this value $\tilde{x}$. Use the simulations in b) to simulate from the *posterior distribution of* $\tilde{x}$. [Hint: the regression curve is a quadratic. You can find a simple formula for $\tilde{x}$ given β_0, β_1 and β_2.]

Problem 3: Monte Carlo

a) Short questions:
   a. Comment on the appropriateness and differences of the following methods 1. Rejection sampling, 2. Metropolis method, 3. Gibbs sampling
      i. when applied in high-dimensional problems
      ii. regarding the number of parameters that need to be adjusted
   b. List all the metrics used to estimate quality of MCMC. Why would one want to generate multiple chains?
   c. Compare5sampling algorithms provided by Turing regarding their speed of convergence, the way they are implemented and other metrics. When would you use each of them?

b) Consider the model

$\eta \mid \theta \sim Binomial(n, \theta),\ \theta \sim Beta(a, b),$

Derive the joint distribution of (η, θ) and the corresponding full conditional distributions. Implement a Gibbs sampler in Julia Turing associated with those full conditionals and compare the outcome (by drawing the empirical distribution obtained by Gibbs sampling against the analytically obtained distribution) of the Gibbs sampler on θ with the true marginal distribution of θ. Assume that n=18, a=b=2.5 and that N=10^5.

Problem 4: More difficult

Read the paper Bayesian Learning via Stochastic Gradient Langevin Dynamics https://www.ics.uci.edu/~welling/publications/papers/stoclangevin_v6.pdf . Reproduce the results of Section 5.1-> Simple Experiment Figures 1 and 2 by implementing your own solution of Stochastic Gradient Langevin Dynamics in Julia.