Markov chain sampling

Week3-ex1, problem statement

The purpose of this exercise is to study the properties of Markov chains and how they can be used to produce samples for Monte Carlo estimation.

Consider a Markov chain defined as follows:

- set $\theta^{(0)} = C$, where C is some constant number.
- for $i = 1, \ldots$ sample $\theta^{(i)} \sim N(\phi\theta^{(i-1)}, \sigma^2)$ where $\phi \in [0, 1)$ is a parameter controlling the autocorrelation between samples.

It can be shown that as $i \to \infty$ (that is, at the limit of long chain) the *marginal* distribution of each $\theta^{(i)}$ is a Gaussian with mean zero and variance $\mathrm{Var}[\theta^{(i)}] = \frac{\sigma^2}{1-\phi^2}$ and that the correlation between $\theta^{(i)}$ and $\theta^{(i+t)}$ is $\mathrm{Corr}([\theta^{(i)}, \theta^{(i+t)}] = \phi^t$

1. What are the variance of $\theta^{(i)}$ and the correlation between $\theta^{(i)}$ and $\theta^{(i+t)}$ at the limit of large i when $\phi \to 0$ and when $\phi \to 1$. Assume that σ is fixed in these cases.
2. Fill in the below table

$\mathrm{Var}[\theta^{(i)}]$	ϕ	σ^2	$\mathrm{Corr}[\theta^{(i)}, \theta^{(i+1)}]$
1		1	
1	0.5		
1		0.2	
1	0.1		

3. Implement the above Markov chain with R and use it to sample random realizations of $\theta^{(i)}$ where $i = 1, \ldots, 100$ with the parameter values given in the above table. As an initial value use $C = 10$. Plot the sample chain and based on the visual inspection, what can you say about the convergence and mixing properties of the chain with the different choices of ϕ?
4. Choose the parameter combination where $\sigma^2 = 0.2$ from the above table. Run three Markov chains with initial values $C_1 = 10$, $C_2 = -10$ and $C_3 = 5$. Find a burn-in value at which the chains have converged according to the PSRF ($\hat{R}$) statistics. This is implemented in function `Rhat` in RStan. Note, $m = 100$ samples might not be enough here.

Note! This is a Markov chain that is constructed very differently from how Stan constructs the Markov chains to sample from the posterior distributions. However, the properties related to autocorrelation and initial value are analogous.

Grading:

Total 10 points: 2 points for correct answer for step 1. 3 points for correct answer to step 2. 3 points for correct answer for step 3. 2 points for correct answer for step 4. **Note**, You should not penalize from wrong parameter values in step 3 and 4 if the table was filled wrong in step 2.

Censored observations

Week3-ex2, problem statement

Exercise instructions

This is the same exercise as 2.1 except that now the posterior inference is performed with MCMC using Stan. Hence, instead of the grid based approximation we use Monte Carlo approximation to do the same analyses as in exercise 2.1.

Suppose you have a Gamma($\alpha = 1, \beta = 1$) prior distribution on the parameter λ which corresponds to the expected number of ship ice besetting events (=events where a ship gets stuck in ice) during 1000 nautical miles in ice infested waters. The number of besetting events, y per distance d (nm) is modeled with a Poisson distribution Poisson($\lambda \times d$). The hyper-parameter α is the shape and β is the inverse scale parameter. You are told that during winters 2013-2017 category A ice breakers traveled in total 6560 nautical miles in the Kara Sea (a sea area in the Arctic Sea). Within this distance they experienced in total more than 2 but less than 7 ice besetting events.

1) Implement the model with Stan and sample from the posterior of λ.
   a) Check for convergence visually and by calculating the PSRF statistics.
   b) Calculate the autocorrelation of the samples.
2) Using the samples of λ
   a) draw the posterior density function of λ and
   b) calculate the posterior probability that $\lambda < 1$ and the 5% and the 95% quantiles.
   c) calculate the posterior mean and variance of λ.
3) Draw samples from the posterior predictive distribution for new $\tilde{y}$ for a ship traveling 1500 nm distance and
   a) draw histogram of samples from the posterior predictive distribution for $\tilde{y}$
   b) Calculate the posterior predictive mean and variance of $\tilde{y}$

Grading

Total 10 points: 4 points for correclty doing step 1. 3 points for correctly doing step 2. 3 points for correctly doing step 3.

Newcomb's speed of light

Week3-ex3, problem statement

R-template `ex_speed_of_light.Rmd`.

Data file `ex_speedOfLight.dat`.

(Here we redo the analysis from page 66 in BDA3.)

Simon Newcomb conducted experiments on speed of light in 1882. He measured the time required for light to travel a certain distance and here we will analyze a data recorded as deviations from 24,800 nanoseconds. The model used in BDA3 is %

$$
\begin{aligned}
y_i &\sim N(\mu, \sigma^2) \\
p(\mu, \sigma^2) &\propto \sigma^{-2}.
\end{aligned}
$$

% where y_i is the i'th measurement, μ is the mean of the measurement and σ^2 the variance of the measurements. Notice that this prior is improper ("uninformative"). This corresponds to widely used uniform prior for μ in the range $(-\infty, \infty)$, and uniform prior for $\log(\sigma)$ (BDA3 pp. 66, 52, and 21). Both priors are improper and cannot be found from Stan. You can use instead %

$$
\begin{aligned}
p(\mu) &\sim N(0, (10^3)^2) \\
p(\sigma^2) &\sim \text{Inv-}\chi^2(\nu = 4, s^2 = 1) \qquad (1)
\end{aligned}
$$

In this exercise your tasks are the following:

1. Write a Stan model for the above model and sample from the posterior of the parameters. Report the posterior mean, variance and 95% central credible interval for μ and σ^2.
2. Additionally draw samples from the posterior predictive distribution of hypothetical new measurement $p(\tilde{y}|y)$. Calculate the mean, variance and 95% quantile of the posterior predictive distribution.
3. How does the posterior predictive distribution differ from the posterior of μ and Why?
4. Which parts of the model could be interpreted to correspond to aleatory and epistemic uncertainty? Discuss whether this distinction is useful here.
5. Instead of Inverse-χ^2 distribution the variance parameter prior has traditionally been defined using Gamma distribution for the precision parameter $\tau = 1/\sigma^2$. By using the results in Appendix A of BDA3 derive the analytic form of a Gamma prior for the precision corresponding to the prior (1). This should be of the form Gamma(α, β), where α and β are functions of ν and s^2.

Note! Many common distributions have multiple parameterizations, for which reason you need to be careful when interpreting others' works. The variance/precision parameter and their priors are notorious for this. The reason is mainly historical since different parameterizations correspond to different analytical solutions.

Grading

Total 10 points. 2 points from correct answer for each of the above steps.