

Computer Lab 4

You are recommended to use R for solving the labs.

You work and submit your labs in pairs, but both of you should contribute equally. and understand all parts of your solutions.

It is not allowed to share exact solutions with other student pairs.

The submitted lab reports will be verified through URKUND and indications of plagiarism will be investigated by the Disciplinary Board.

Submit your solutions via LISAM, no later than May 26 at 23:30.

1. Time series models in Stan

- (a) Write a function in R that simulates data from the AR(1)-process

$$x_t = \mu + \phi(x_{t-1} - \mu) + \varepsilon_t, \quad \varepsilon_t \stackrel{iid}{\sim} N(0, \sigma^2),$$

for given values of μ , ϕ and σ^2 . Start the process at $x_1 = \mu$ and then simulate values for x_t for $t = 2, 3, \dots, T$ and return the vector $x_{1:T}$ containing all time points. Use $\mu = 10$, $\sigma^2 = 2$ and $T = 200$ and look at some different realizations (simulations) of $x_{1:T}$ for values of ϕ between -1 and 1 (this is the interval of ϕ where the AR(1)-process is stable). Include a plot of at least one realization in the report. What effect does the value of ϕ have on $x_{1:T}$?

- (b) Use your function from a) to simulate two AR(1)-processes, $x_{1:T}$ with $\phi = 0.3$ and $y_{1:T}$ with $\phi = 0.95$. Now, treat your simulated vectors as synthetic data, and treat the values of μ , ϕ and σ^2 as unknown and estimate them using MCMC. Implement Stan-code that samples from the posterior of the three parameters, using suitable non-informative priors of your choice. [Hint: Look at the time-series models examples in the Stan user's guide/reference manual, and note the different parameterization used here.]
- Report the posterior mean, 95% credible intervals and the number of effective posterior samples for the three inferred parameters for each of the simulated AR(1)-process. Are you able to estimate the true values?
 - For each of the two data sets, evaluate the convergence of the samplers and plot the joint posterior of μ and ϕ . Comments?
- (c) The data `campy.dat` contain the number of cases of campylobacter infections in the north of the province Quebec (Canada) in four week intervals from January 1990 to the end of October 2000. It has 13 observations per year and 140 observations in total. Assume that the number of infections c_t at each time point follows an independent Poisson distribution when conditioned on a latent AR(1)-process x_t , that is

$$c_t | x_t \sim \text{Poisson}(\exp(x_t)),$$

where x_t is an AR(1)-process as in a). Implement and estimate the model in Stan, using suitable priors of your choice. Produce a plot that contains both the data and the posterior mean and 95% credible intervals for the latent intensity $\theta_t = \exp(x_t)$ over time. [Hint: Should x_t be seen as data or parameters?]

- (d) Now, assume that we have a prior belief that the true underlying intensity θ_t varies more smoothly than the data suggests. Change the prior for σ^2 so that it becomes informative about that the AR(1)-process increments ε_t should be small. Re-estimate the model using Stan with the new prior and produce the same plot as in c). Has the posterior for θ_t changed?

HAVE FUN!