Project exam for STK4051/9051 - Computational statistics

Spring 2019

Part 2 (of 2)

This is the second part of the compulsory exercise for STK4051/9051, spring semester 2019. The deadline for the complete compulsory exercise (including part 1) is April 30. This must be delivered in the Devilry system (`devilry.ifi.uio.no`).

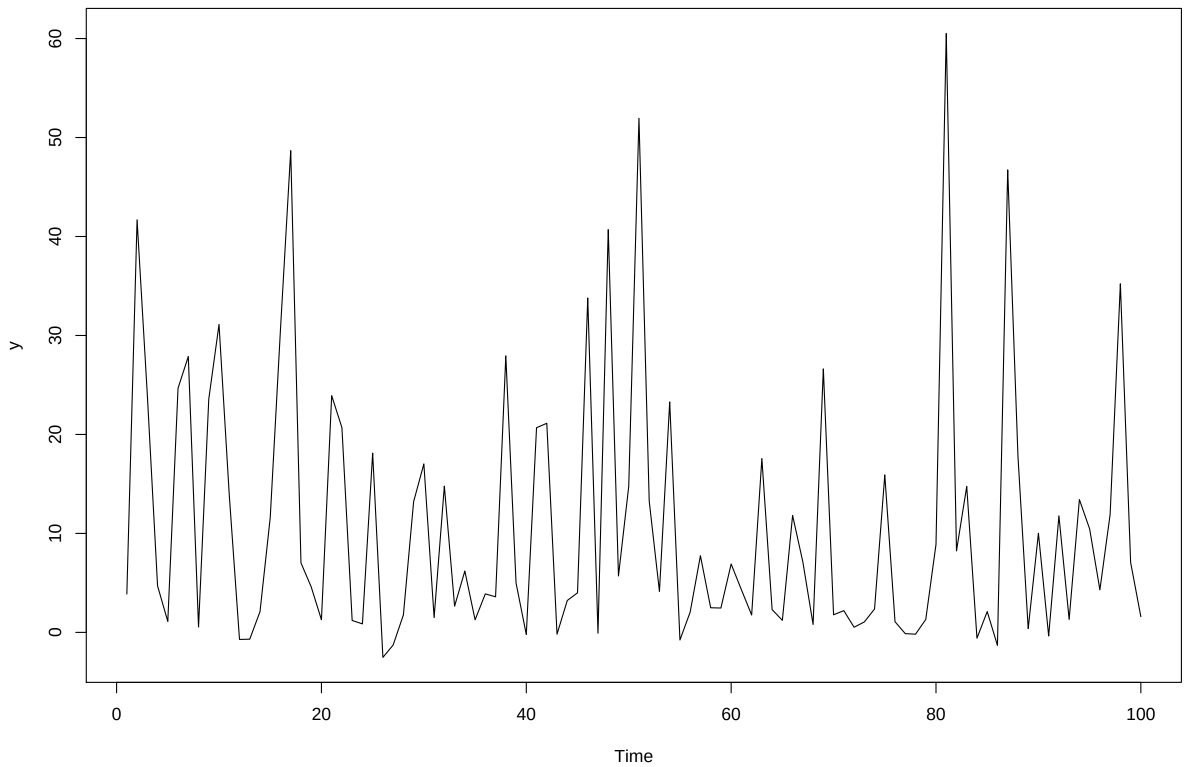
Reports may be written in Norwegian or English, and should preferably be text-processed (LaTeX, Word). Give your name and "student number" on the first page. Write concisely. Relevant figures need to be included in the report. Copies of relevant parts of machine programs used (in R, or matlab, or similar) are also to be included, perhaps as an appendix to the report.

This set contains two exercises and comprises three pages. Data is available from the course web-page. Within these exercises there is some choices you need to make when designing algorithms. Part of the evaluation will be on your choices, but more importantly is your arguments on why you have made the specific choices!

Exercise 1 (Sequential Monte Carlo). We will in this exercise consider inference of the dynamic structure based on a constructed model:

$$
\begin{aligned}
x_1 \sim& N(0,10)\\
x_t =& 0.5 * x_{t-1} + 25 \cdot x_{t-1}/(1+x_{t-1}^2) + 8 \cdot \cos(1.2*(t-1)) + \varepsilon_t, \qquad && t = 2,3,..\\
y_t =& x_t^2/25 + \eta_t, && t = 1,2,3,...
\end{aligned}
$$

where all the noise terms $\{\varepsilon_t\}$ and $\{\eta_t\}$ are independent and $\varepsilon_t \sim N(0,\sigma_x^2)$ while $\eta_t \sim N(0,\sigma_y^2)$. The plot below show data $\{y_t\}$ simulated from this model with $\sigma_x^2 = 10$ and $\sigma_y^2 = 1$. The data is available at the course web-page: *https://www.uio.no/studier/emner/matnat/math/STK4051/h17/data/stdr_example_y.txt*



The aim of the exercise will be to make inference on the latent process $\{x_t\}$. We will first assume σ_x^2 and σ_y^2 are known.

(a). We will start by considering simulation from $p(x_1|y_1)$. Discuss different methods for obtaining properly weighted samples for estimation of $E[h(x_1)|y_1)]$.

Perform such simulations and use this to obtain estimates and 95% uncertainty intervals for x_1 and x_1^2 conditional on y_1. Discuss these results.

(b). Design a sequential Monte Carlo algorithm for inference about $p(x_t|\boldsymbol{y}_{1:t})$ for $t = 1, 2, ...$ Indicate clearly what choices you make when designing your algorithm.

(*c*). Run the algorithm you designed in (*b*) for the given data. Use this to make inference on x_t and x_t^2 based on data up to time t.

Discuss the results.

(*d*). Also use your output to make inference about x_t and x_t^2 based on all the data. Discuss potential problems with this.

(*e*). Assume now σ_x^2 is unknown but has a prior $\tau_x \sim \text{Gamma}(0.01, 0.01)$ where $\tau_x = 1/\sigma_x^2$. Construct a sequential Monte Carlo method for estimation of σ_x^2 (or τ_x) as well. Again, make clear what choices you make when designing the algorithm. Discuss the performance of your method.

Exercise 2 (Markov chain Monte Carlo). Consider again the same model and data as in the first exercise. Assume again first that σ_x^2 and σ_y^2 are known. An alternative procedure for making inference about $\{x_t\}$ is to construct an MCMC algorithm for simulation from $p(\boldsymbol{x}_{1:T}|\boldsymbol{y}_{1:T})$ where $T = 100$ is the number of observations.

(*a*). Show that if $\boldsymbol{x}_{1:T}$ and $\boldsymbol{x}^*_{1:T}$ only differ in the t-th component, then

$$\frac{p(\boldsymbol{x}^*_{1:T}|\boldsymbol{y}_{1:T})}{p(\boldsymbol{x}_{1:T}|\boldsymbol{y}_{1:T})}$$

only depend on x_{t-1}, x_t, x_t^* and x_{t+1} (with some modifications in the end-points).

(*b*). Design a Metropolis-Hasting algorithm for simulation from $p(\boldsymbol{x}_{1:T}|\boldsymbol{y}_{1:T})$. Indicate clearly what choices you make when designing your algorithm.

(*c*). Run the algorithm in (*b*) on the given data. Discuss the results you obtained compared to the ones from Exercise 1.

(*d*). Assume now σ_x^2 is unknown but has a prior $\tau_x \sim \text{Gamma}(0.01, 0.01)$ where $\tau_x = 1/\sigma_x^2$. Derive the distribution for $p(\tau_x|\boldsymbol{x}_{1:T}, \boldsymbol{y}_{1:T})$. Use this to construct an algorithm for simulation from $p(\boldsymbol{x}_{1:T}, \tau_x|\boldsymbol{y}_{1:T})$.

(*e*). Run the algorithm in (*d*) on the given data. Discuss the results you obtained compared to the ones from Exercise 1.