Assignments: Multivariate Simulation

Niels Richard Hansen

September 11, 2017

The last assignment topic is multivariate simulation. If you draw the topic "Multivariate simulation" at the oral exam, you will have to present a solution of one of the two assignments below.

Remember the five points:

- How can you test that your implementation is correct?
- Can you implement alternative solutions?
- Can the code be restructured e.g. by modularization, abstraction or object oriented programming to improve readability?
- How does the implementation perform (benchmarking)?
- Where are the bottlenecks (profiling), and what can you do about them?

For both of the assignments below the core problem is simulation of a Markov chain. This is done via an update formula, but R is not an ideal language for implementation of such sequential algorithms due to their iterative nature. The assignments are thus suitable for experimenting with Rcpp. However, for a complete solution, it is not a requirement that you provide an Rcpp implementation. You can also explore benefits and deficits with various implementations in R.

It should be possible to modify R code from the lectures relatively easily to obtain a solution for both assignments. You are welcome to do so, but you should also explore other possibilities. That is, don't just modify my code in the obvious places. Extend my code with additional functionality. Improve my code in terms of performance, modularity, and robustness.

Assignment 1: Sequential Monte Carlo

The framework of this assignment is Section 6.3.2.4. The model you will consider is given by the update equations

- $\varepsilon_t \sim \mathcal{N}(0, \sigma^2)$
- $X_t = \max(\alpha 1(X_{t-1} \leq \gamma) + \beta 1(X_{t-1} > \gamma) + \varepsilon_t, 1)$
- $Y_t \sim \text{Pois}(X_t)$

for parameters $\alpha, \beta \in \mathbb{R}$ and $\sigma > 0$ and with $\gamma = (\alpha + \beta)/2$. You can assume throughout that $X_0 > 0$.

The model is a Poisson model of Y_t, whose mean, X_t, will change with time according to how this process evolves. This can be used to model events that happen in a heterogeneous and clustered way over time. There are many usages, one could be traffic modelling, another insurance claims and a third the spike times of a neuron in the brain. Common to all such applications is that X_t is not observed. The prediction of X_t from the observation of $Y_0, \ldots, Y_t$ is what we call the filtration problem.

The purpose of this assignment is to implement a filter for this model using sequential Monte Carlo methods. Implement first an algorithm for simulation from the model. Implement then one or more filters (there are different options in terms of how often resampling is done). Investigate the performance of the implementation(s) using simulations, and apply the implementation to the SMC data (see Absalon). The data set contains 500 observations of the Y_ts for the model with $\alpha = 5$, $\beta = 20$ and $\sigma = 3$. You can assume that $X_0 = 1$.

It is a good idea to focus on computing $E(X_t \mid Y_0, \ldots, Y_t)$. Plot this conditional expectation as a function of t and compare with the observations.

Assignment 2: Markov Chain Monte Carlo

Solve Problem 7.8 a+b in CS (skip Question c).

The theoretical setup for Problem 7.8 is Problem 7.7. You will need the results, but you won't need to present the solution of Problem 7.7. It is unclear what the index notation $j(i)$ actually means in the formulas from Problem 7.7. I cannot make anything else of it than that it means batch i and sample j, and that it could (should?) have been written ij. You should note that if σ_ϵ^2 and σ_β^2 were not known, it would be impossible to identify them from the averaged model (7.30). One could estimate σ_ϵ^2 by computing the variance of the difference between the two replications (not available in the data set) for each combination of batch and sample. In this sense it can be partly justified to assume σ_ϵ^2 known. The other two variances would in practice have to be either estimated or given prior distributions as well.

The main purpose of this assignment is to give you some experience with implementing and running a Gibbs sampler. This is an important workhorse used for Bayesian analysis of hierachical models as implemented in the classical WinBUGS and now OpenBUGS programs. The R package BRugs gives an R interface. It does not run directly on Mac. More recent software for Bayesian modeling is Stan.