Hand-in assignments
PhD course on Sequential Monte Carlo methods 2021

Linköping University and Uppsala University

August 26, 2021

Deadline (in order for to get your course certificate on time): 30 September 2021.

Instructions: Hand in your solutions (including plots etc) to *all* problems in this document compiled as *one* PDF to `johan.alenlov@liu.se`, with topic `SMC2021:Homework`. You should also append your implementation (written in a suitable language of your choice, e.g., Matlab, R, Python or Julia) as a zip-file (or similar) to the e-mail. There should, however, be no need to read the code in order to grade your solution, i.e. the pdf should contain all important derivations and results, as well as important design choices made in the implementation. (You should not hand in solutions to the exercise problems in the documents named "Exercises".)

H.1 Importance sampling theory

(a) Consider importance sampling with proposal $q(x)$ and target $\pi(x) = \frac{\widetilde{\pi}(x)}{Z}$. Show that $\widehat{Z} = \frac{1}{N}\sum_{i=1}^{N} \frac{\widetilde{\pi}(X^i)}{q(X^i)}$ (where X^i are independent random variables distributed according to q) is an unbiased estimator of the normalizing constant Z. [2p]

(b) Under appropriate conditions, the importance sampling normalization constant estimator will satisfy a central limit theorem,

$$\sqrt{N}\left(\frac{\widehat{Z}}{Z} - 1\right) \xrightarrow{\mathrm{d}} \mathcal{N}\left(0, \int \frac{\pi(x)^2}{q(x)}\mathrm{d}x - 1\right).$$

Consider importance sampling with target $\pi(x) = \mathcal{N}(x \,|\, 0, 1)$ and proposal $q(x) = \mathcal{N}(x \,|\, 0, \lambda^{-1})$. What is the condition on λ for the asymptotic variance of the normalization constant estimator to be finite? Interpret the result. [3p]

H.2 Particle filter for a linear Gaussian state-space model

Consider the following scalar linear Gaussian state-space model

$$X_t = 0.9X_{t-1} + V_t, \qquad V_t \sim \mathcal{N}(0, 0.5), \tag{1a}$$
$$Y_t = 1.3X_t + E_t, \qquad E_t \sim \mathcal{N}(0, 0.1). \tag{1b}$$

Let the initial state be distributed according to $X_0 \sim \mathcal{N}(0, 1)$.

(a) Write this model in the form

$$X_t \,|\, (X_{t-1} = x_{t-1}) \sim p(x_t \,|\, x_{t-1}), \tag{2a}$$
$$Y_t \,|\, (X_t = x_t) \sim p(y_t \,|\, x_t). \tag{2b}$$

In other words, find the probability density functions $p(x_t \,|\, x_{t-1})$ and $p(y_t \,|\, x_t)$ in (2) corresponding to the model (1). Simulate the model to produce $T = 2\,000$ synthetic measurements $y_{1:T}$. [1p]

(b) Since this is a linear Gaussian state-space model, an analytical solution for $p(x_t \,|\, y_{1:t})$ $(t = 1, \ldots, T)$ exists

and is given by the Kalman filter. The Kalman filter for a model

$$X_t = AX_{t-1} + V_t, \qquad V_t \sim \mathcal{N}(0, Q), \tag{3a}$$
$$Y_t = CX_t + E_t, \qquad E_t \sim \mathcal{N}(0, R), \tag{3b}$$
$$X_0 \sim \mathcal{N}(0, P_0). \tag{3c}$$

is given by the recursions

$$\widehat{x}_{t\,|\,t} = A\widehat{x}_{t-1\,|\,t-1} + K_t(y_t - CA\widehat{x}_{t-1\,|\,t-1}), \tag{4a}$$
$$\widehat{P}_{t\,|\,t} = \widehat{P}_{t\,|\,t-1} - K_t C\widehat{P}_{t\,|\,t-1}, \tag{4b}$$

with

$$\widehat{P}_{t\,|\,t-1} = A\widehat{P}_{t-1\,|\,t-1}A^T + Q, \tag{4c}$$
$$K_t = \widehat{P}_{t\,|\,t-1}C^T(C\widehat{P}_{t\,|\,t-1}C^T + R)^{-1}, \tag{4d}$$
$$\widehat{P}_{0\,|\,0} = P_0, \tag{4e}$$
$$\widehat{x}_{0\,|\,0} = 0, \tag{4f}$$

from which the filtering density is given by $p(x_t \,|\, y_{1:t}) = \mathcal{N}(x_t \,|\, \widehat{x}_{t\,|\,t}, \widehat{P}_{t\,|\,t})$. Implement this recursion, and compare the mean of $p(x_t \,|\, y_{1:t})$ against the "true" state trajectory (which is available since you have simulated the data yourself) in a plot. Can you in general expect the mean of the particle filter estimate of $p(x_t \,|\, y_{1:t})$ to be closer to the "true" trajectory than the Kalman filter? Why/why not? [2p]

(c) Implement the bootstrap particle filter (with multinomial resampling) and compare its mean and variance estimates to the Kalman filter results. (Compare the mean and variance of $p(x_t \,|\, y_{1:t})$ for each time step $t = 1, \ldots, T$, and report the average absolute difference.) Report the results for $N = 10, 50, 100, 2000, 5000$ numbers of particles in the form of a table. [4p]

Hint: For a correct particle filter implementation, you can expect the error to vanish as $N \to \infty$.

(d) Derive the expressions for the locally optimal proposals

$$\nu_{t-1}^i = p(y_t \,|\, x_{t-1}^i), \quad i = 1, \ldots, N, \qquad \text{(resampling weights)}$$
$$q(x_t \,|\, x_{t-1}, y_t) = p(x_t \,|\, x_{t-1}, y_t), \qquad \text{(propagation proposal)}$$

for the model given in (1). Implement the fully adapted particle filter and run it on the same simulated data as above. Compute the errors compared to the Kalman filter (as you did for the bootstrap particle filter), and compare it to the errors for the bootstrap particle filter (using the same number of particles N). Which method seems better? [3p]

(e) Run the fully adapted particle filter using $N = 100$ particles. Compute the trajectories $\{x_{0:T}^i\}_{i=1}^N$ by tracing the particle genealogy from time T back to time[1] 0. Plot all the $N = 100$ trajectories in one figure. [2p]

(f) Change the resampling method to systematic resampling, rerun the method and plot the particle genealogy again. Do you see any changes, and if so, why? [2p]

(g) Add adaptive resampling with an ESS trigger of $N/2 = 50$. Rerun the method and plot the particle genealogy and N_{ESS}/N as a function of time. Comment on the result. [2p]

[1] The plot might be hard to read if you plot from 0 to $T = 2\,000$. If you prefer, you can plot only a certain interval, as long as the path degeneracy situation is clear from your plot.

H.3 Parameter estimation in the stochastic volatility model

Consider the so-called stochastic volatility model

$$X_t|(X_{t-1} = x_{t-1}) \sim \mathcal{N}(x_t \,|\, \phi x_{t-1}, \sigma^2), \tag{5a}$$

$$Y_t|(X_t = x_t) \sim \mathcal{N}(y_t \,|\, 0, \beta^2 \exp(x_t)), \tag{5b}$$

where the parameter vector is given by $\theta = \{\phi, \sigma, \beta\}$. Here, X_t denotes the underlying latent volatility (the variations in the asset price) and Y_t denotes the observed scaled log-returns from some financial asset. The $T = 500$ observations that we consider in this task are log-returns from the NASDAQ OMX Stockholm 30 Index during a two year period between January 2, 2012 and January 2, 2014. We have calculated the log-returns by $y_t = 100[\log(s_t) - \log(s_{t-1})]$, where s_t denotes the closing price of the index at day t. The data is found in `seOMXlogreturns2012to2014.csv`. For more details on stochastic volatility models, see e.g. [1, 2].

(a) Let ϕ be unknown, and assume the other parameters are $\sigma = 0.16$ and $\beta = 0.70$. Make a reasonably coarse grid for ϕ between 0 and 1, and use the bootstrap particle filter to estimate the log-likelihood for each of these values of ϕ. Compute 10 estimates for each grid point and plot all of them in one figure as, e.g., boxplots (ϕ on the horizontal axis, log-likelihood estimates on the vertical axis). [3p]

(b) Assume now instead that the variance parameters σ^2 and β^2 are unknown, whereas $\phi = 0.985$ is assumed to be known. We consider a Bayesian setting and place inverse Gamma priors on the parameters:

$$\sigma^2 \sim \mathcal{IG}(a = 0.01, b = 0.01), \tag{6a}$$

$$\beta^2 \sim \mathcal{IG}(a = 0.01, b = 0.01), \tag{6b}$$

where the inverse Gamma pdf with parameters (a, b) is given by

$$\mathcal{IG}(x \,|\, a, b) = \frac{b^a}{\Gamma(a)} x^{-a-1} \exp\left(-\frac{b}{x}\right) \tag{7}$$

and where $\Gamma(\cdot)$ is the Gamma function.

Implement the PMH algorithm to compute the posterior distribution $p(\sigma^2, \beta^2 \,|\, y_{1:T})$, using a Gaussian random walk proposal. Report the marginal distributions (e.g., histograms) for the two parameters. [4p]

Hint: The priors are only supported on the positive real line. Thus, if a negative value is proposed for any of the two parameters, this can be rejected without having to compute a likelihood estimate. There are better ways of handling parameter constraints, but for simplicity we will not consider them here.

Note: Be aware that the parametrization of the Gamma and inverse Gamma distribution is not completely standardized, and different software packages use different conventions.

Note: The priors are rather vague (in fact, they do not even have a finite variance), and samples from the prior can take quite large values. Not, however, that you never have to sample from the prior!

(c) Implement a particle Gibbs sampler for the same task as in the previous problem. Note that the inverse Gamma prior is *conjugate* to the model (5). Thus, the full conditional distributions for the parameters are available in closed form:

$$p(\sigma^2 \,|\, x_{0:T}, y_{1:T}) = p(\sigma^2 \,|\, x_{0:T}) = \mathcal{IG}\left(\sigma^2 \,|\, a + \frac{T}{2}, b + \frac{1}{2}\sum_{t=1}^{T}(x_t - \phi x_{t-1})^2\right), \tag{8a}$$

$$p(\beta^2 \,|\, x_{0:T}, y_{1:T}) = \mathcal{IG}\left(\beta^2 \,|\, a + \frac{T}{2}, b + \frac{1}{2}\sum_{t=1}^{T}\exp(-x_t)y_t^2\right) \tag{8b}$$

[4p]

Hint: Both PMH and particle Gibbs are consistent MCMC samplers. Thus, as a sanity check you should make sure that both algorithms find the same posterior distributions if run for a large number of MCMC iterations.

Hint: Don't forget to discard the transient phase (the "burn-in") of the MCMC chains.

H.4 **SMC sampler**

Consider a pdf[2] in $\mathbb{R}^2$,

$$\pi(x) = \frac{1}{Z}\mathbb{1}(-1 \leq x_1 \leq 1)\mathbb{1}(-1 \leq x_2 \leq 1)\sin^2(x_1\pi)\cos^8(x_2 2\pi)\exp(-5(x_1^2 + x_2^2)) \tag{9}$$

where $x = (x_1, x_2)$. Implement an SMC sampler to sample from $\pi(x)$ and to estimate the normalizing constant Z. Use $N = 100$ particles and an appropriately chosen annealing sequence of K steps, such that it is possible to sample exactly from the initial distribution in the sequence. Use ESS-triggered resampling (threshold $0.7N$) and select the annealing sequence in such a way that the average number of SMC iterations between consecutive resampling steps is in the order of 10. As a proposal in Metropolis-Hastings, use a Gaussian random walk independent in each dimension with variance 0.02^2, or similar. Report the results by:

(a) Plotting the locations of the particles at initialization, at the final iteration, and at two arbitrarily chosen intermediate iterations (or, if you prefer, make an animation over k). [4p]

(b) Plot the ESS as a function of iteration number k. What happens with the ESS when a resampling is triggered? [2p]

(c) Reporting the estimate of the normalizing constant Z. [2p]

References

[1] Marc Chesney and Louis Scottt. "Pricing European currency options: A comparison of the modified Black-Scholes model and a random variance model". In: *Journal of financial and quantitative analysis* 24.3 (1989), pp. 267–284.

[2] Angelo Melino and Stuart M. Turnbull. "Pricing foreign currency options with stochastic volatility". In: *Journal of Econometrics* 45.1–2 (1990), pp. 239–265.

[2]The symbol π used in the argument to the trigonometric functions is the constant $3.14\ldots$, not the target distribution.