Exercises set I
PhD course on Sequential Monte Carlo methods 2021

Linköping University and Uppsala University

August 16, 2021

This document contains exercises to make you familiar with the content of the course. *The exercises in this document are not mandatory, and you do not need to hand in your solutions.* The mandatory assignment is found in a separate document named "Hand-in". We strongly recommend that you carefully work through these exercises before starting with the mandatory assignments.

I.1 **Importance sampling.**

Let us for this problem assume that you can only generate random numbers with a standard normal distribution, $q(x) = \mathcal{N}(x; 0, 1)$, but are interested in (possibly weighted) samples from $\pi(x) = \mathcal{U}(x; 0, 4)$. $\mathcal{N}$ denotes the normal (Gaussian) distribution, and $\mathcal{U}$ the uniform distribution.

(a) Consider an importance sampler with proposal $q(x) = \mathcal{N}(x; 0, 1)$ and target $\pi(x)$. Is this a valid importance sampler?

(b) Implement the suggested importance sampler with $N = 10\,000$. Plot the result as, for example, a kernel density estimate or a weighted histogram. What problems do you experience with your sampler, and how can you improve it?

(c) For importance sampling it holds that an estimate of any test function ϕ is unbiased. Use the weighted samples to estimate the mean of the target ($\phi(x)$ is the identity function), and make a simulation study to support that this estimate is indeed unbiased (regardless of choice of proposal q). *Note that this claim holds for any finite number of samples N!* Confirm this by using only a small number of samples, e.g., $N = 10$, in your simulations.

(d) We have so far assumed that we can evaluate $\pi(x)$ exactly. However, sometimes we can only evaluate the target $\pi(x)$ up to proportionality, as $\pi(x) = \frac{\widetilde{\pi}(x)}{Z}$, where we can evaluate $\widetilde{\pi}(x)$ but the normalizing constant Z is unknown to us. Actually, sometimes Z is the quantity of interest. (If, e.g., the target $\pi(x)$ is a posterior, then $\widetilde{\pi}(x)$ is the likelihood times the prior, and Z is the marginal likelihood which can be useful for, e.g., model selection.)

Give an informal derivation to the estimator

$$\widehat{Z} = \frac{1}{N}\sum_{i=1}^{N} \widetilde{W}^i, \text{ where } \widetilde{W}^i = \frac{\widetilde{\pi}(X_i)}{q(X_i)}.$$

Hint: Start with $Z = \int \widetilde{\pi}(x)dx$.

(e) Implement the estimator to estimate Z if $\widetilde{\pi}$ is the indicator function for the interval $[0, 4]$. Make a simulation study supporting the theoretical claim that $\widehat{Z}$ is an unbiased estimate of Z (for some finite N). Also explore how the variance of $\widehat{Z}$ changes with more or less 'good' choices of proposals q.

(f) The estimator $\widehat{Z}$ is indeed unbiased, but we still have to take care when using it. A typical case is when we can only evaluate π up to proportionality (i.e., $\widetilde{\pi}$), but are still interested in functionals of π. (Note that Z is actually a functional of $\widetilde{\pi}$ rather than π.) Let us say that we are once again (cf. (c)) interested in estimating the mean of the target, but we can only access $\widetilde{\pi}$, which here is the indicator function for the interval $[0, 4]$. Derive your estimator for the mean of π for this case, and confirm with a simulation study that it is *not* unbiased. (Note: your proposal q has to be asymmetric around the true mean in order to see the effect. Use, e.g. $q(x) = \mathcal{N}(x; 0, 3^2)$ and a small number of samples, e.g., $N = 10$.) Also confirm that the bias vanishes as $N \to \infty$.

(g) Show that the solution to the previous problem essentially corresponds to simply normalizing the sample weights w_i such that $\sum_i w_i = 1$ before reporting the mean estimate.

I.2 **Importance sampling in higher dimensions.**

Let us consider importance sampling in a D-dimensional space. Assume that you have access to D-dimensional random vectors from the standard multivariate normal distribution $\mathcal{N}(0, I_D)$ (I_D is here the D-dimensional identity matrix), and use importance sampling to generate weighted samples from the D-dimensional multivariate uniform distribution $\pi(x) = \mathcal{U}(x; [-.5, .5]^D)$, i.e., the unit cube centered around the origin. (In this problem, you can evaluate the target $\pi(x)$ exactly.) Repeat the experiment for different values of D, and plot the proportion of samples with non-zero weights as a function of D. Conclude at what rate this fraction decreases (constant, linear, polynomial, exponential, etc.) based on both your simulations as well as a theoretical argument.

I.3 **An important numerical aspect**

Consider again the high-dimensional problem, but this time with focus on the numerical aspects in $D = 1\,000$ dimensions:

(a) Consider importance sampling in $\mathbb{R}^D$, with target $\widetilde{\pi}(x) \propto \mathcal{N}(x; 0, I_D)$ (i.e., x is a D-dimensional vector). Use the proposal $q(x) = \mathcal{N}(x; 0, 2^2 \cdot I_D)$. Generate $N = 10$ samples x^i and compute, for each sample, first $\widetilde{\pi}(x^i)$ and then $q(x^i)$. Note that since the covariance matrices are diagonal, the densities factorize $\mathcal{N}(x; 0, I_D) = \prod_{k=1}^{D} \mathcal{N}(x_k; 0, 1)$ etc, where x_k is the kth component of x. Next, compute the weights $\widetilde{w}^i = \frac{\widetilde{\pi}(x^i)}{q(x^i)}$ and the normalized version of it $w^i = \frac{\widetilde{w}^i}{\sum_i \widetilde{w}^i}$. What normalized weights w^i do you obtain? Why?

(b) The perhaps most useful remedy to this problem is to consider the logarithm of the weights instead. Hence, use that $\log \mathcal{N}(x; 0, I_D) = \sum_{k=1}^{D} \log \mathcal{N}(x_k; 0, 1)$, and compute $\log \widetilde{w}^i = \log \widetilde{\pi}(x^i) - \log q(x^i)$ instead. Do you experience the same problem?

(c) If $\log \widetilde{w}^i$ is too small, $\exp(\log \widetilde{w}^i)$ may still be smaller than what the system can represent. In order to obtain a normalized version of the weights, explore the 'trick' of computing $\underline{w}^i = \exp(\log \widetilde{w}^i - \max_j \{\log \widetilde{w}^j\})$ instead, and use $\underline{w}^i$ to obtain the normalized weights instead. Why is this a valid approach?

The aspects explored in this problem are highly relevant also for state-space filtering problems when the state dimension is small, but T large. Our recommendation is to always *implement the logarithm of the weights when working with Monte Carlo!*

I.4 **Bootstrap particle filter for the stochastic volatility model**

Consider the so-called stochastic volatility model

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

$$y_t | x_t \sim \mathcal{N}(y_t; 0, \beta^2 \exp(x_t)), \tag{1b}$$

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. [**CS:1989**, **MT:1990**].

Assume that the parameter vector is given by $\theta = \{0.98, 0.16, 0.70\}$. Estimate the marginal filtering distribution at each time index $t = 1, \ldots, T$ using the bootstrap particle filter with $N = 500$ particles. Make a reasonable assumption about the initial state x_0. Plot the mean of the filtering distribution $p(x_t \mid y_{1:t-1})$ at each time step and compare with the observations. Is the estimated volatility reasonable?

I.5 **Bootstrap particle filter central limit theorem**

Recall the bootstrap particle filter central limit theorem,

$$\sqrt{N}\left(\sum_{i=1}^{N} W_t^i \varphi(X_t^i) - I_t(\varphi)\right) \xrightarrow{\mathrm{d}} \mathcal{N}(0, V_t(\varphi))$$

with

$$\begin{aligned} V_t(\varphi) &= \sum_{k=0}^{t} \int \frac{p(x_k \mid y_{1:t})^2}{p(x_k \mid y_{1:k-1})} \left(I_{k,t}(\varphi \mid x_k) - I_t(\varphi)\right)^2 \mathrm{d}x_k, \qquad (2)\\ I_t(\varphi) &= \mathbb{E}[\varphi(X_t) \mid y_{1:t}]\,,\\ I_{k,t}(\varphi \mid x_k) &= \mathbb{E}[\varphi(X_t) \mid y_{k+1:t}, x_k] = \begin{cases} \varphi(x_t), & \text{if } k = t,\\ \int \varphi(x_t) p(x_t \mid x_k, y_{k+1:t}) \mathrm{d}x_t, & \text{if } k < t. \end{cases} \end{aligned}$$

Use the recursive expressions for the asymptotic variance (see lecture 5)

$$\begin{aligned} \widetilde{V}_{t-1}(\varphi) &= V_{t-1}(\varphi) + \mathrm{Var}[\varphi(X_{t-1}) \mid y_{1:t-1}]\,, && \text{(resampling)}\\ \bar{V}_t(\varphi) &= \widetilde{V}_{t-1}(\mathbb{E}[\varphi(X_t) \mid x_{t-1}]) + \mathbb{E}[\mathrm{Var}[\varphi(X_t) \mid X_{t-1}] \mid y_{1:t-1}]\,, && \text{(propagation)}\\ V_t(\varphi) &= \bar{V}_t\left(\frac{p(y_t \mid x_t)}{p(y_t \mid y_{1:t-1})} \cdot \{\varphi(x_t) - \mathbb{E}[\varphi(X_t) \mid y_{1:t}]\}\right), && \text{(weighting)} \end{aligned}$$

to verify the additive expression for the variance (**??**).

Hint: Start by verifying that

$$V_t(\varphi_t) = V_{t-1}(\varphi_{t-1}) + \mathrm{Var}[\zeta_t(X_t) \mid y_{1:t-1}]$$

for some functions $\varphi_{t-1}(x_{t-1})$ and $\zeta_t(x_t)$ which are expressed in terms of $\varphi_t(x_t) = \varphi(x_t)$. Then compute the two terms of the above expression explicitly using an inductive argument.