ADVANCED SIMULATION METHODS

COURSEWORK 1, SPRING 2024-2025

This assignment will be done individually. You should prepare a **PDF** report written primarily in a in a report/paper format, to be submitted through Blackboard, Coursework 1 container (under Courseworks folder). You can also write this in IPython notebook format, if you like, however, you should export a readable PDF from this. By submitting the coursework, you will be agreeing that this assessed coursework is your own work unless otherwise acknowledged, includes no plagiarism, you have not discussed it with anyone else except when seeking clarifications with the module lecturer, and have not shared any code with anyone else prior to submission. In your answers, you can quote material from the notes and shared IPython notebooks, providing your own original explanations.

This course work accounts for approximately 40% of the final grade. The report will be marked from 0 to 40. The deadline is **1PM, 12 March 2025**. To enable anonymous marking, please put your CID and not your name into the filename of the submission or into the manuscript.

Q1: ESTIMATING A LATENT VARIABLE MODEL

Assume that we have the following latent variable model:

$$p(\theta) = \mathcal{N}(\theta; m, P),$$
$$p(x|\theta) = \mathcal{N}(x; C\theta, Q),$$
$$p(y|x) = \mathcal{N}(y; Ax, R),$$

where

$$m = \begin{bmatrix} 2 \\ 0 \end{bmatrix}, \qquad P = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \qquad C = \begin{bmatrix} 1 & 1 \\ 1 & 1.01 \end{bmatrix},$$

$$Q = \begin{bmatrix} 0.01 & 0 \\ 0 & 0.01 \end{bmatrix}, \qquad A = \begin{bmatrix} 1 & 0 \\ \epsilon & 0 \end{bmatrix}, \qquad R = \begin{bmatrix} \epsilon & 0 \\ 0 & \epsilon \end{bmatrix}.$$

Below, we set our observed data to

$$y = \begin{bmatrix} 1 \\ 1 \end{bmatrix}.$$

The parameter ϵ is a free parameter which we will use to investigate sampling behaviour. In this question, we aim at sampling from the posterior $p(\theta|y)$ using various methods you have seen in Chapters 1 and 2.

1. Provide the expressions of the distributions $p(y|\theta)$ and $p(\theta|y)$ in closed form using the Gaussian marginalization and conditioning formulas in Chapter 2. You can refer to the lemmas in the lecture notes, but do give a brief explanation of the derivation. (2 marks)

2. Given the parameters in the main part of the question and your derivation in Part 1, compute (on computer) the mean and covariance of $p(\theta|y)$ and provide a contour plot (see this Jupyter notebook for helper code for plotting) for $\epsilon = 10^{-3}, 10^{-2}, 10^{-1}, 1$. Discuss these plots and the effect of ϵ. (1 mark)

3. Fix $\epsilon = 1$. Present and demonstrate the performance of the random-walk Metropolis-Hastings (RWMH) sampler to sample from $p(\theta|y)$. Provide a clear explanation of the method, the parameter choice (e.g. proposal variance, denoted σ_{MH}^2). Draw $N = 100000$ samples after an appropriate `burnin` stage (you need to set this) to demonstrate the method. (2 marks)

4. Provide the plots of the autocorrelation functions w.r.t. different proposal variance choices by varying σ_{MH}^2 between 10^{-4} to 4 with an appropriate precision and running RWMH for these variances. Identify the best variance for mixing. Plot your samples for the best choice of your proposal on top of the density contour. (3 marks)

5. Fix the variance σ_{MH}^2 of the proposal to the best value you have obtained in the previous part, now try this sampler for $\epsilon = 10^{-3}, 10^{-2}, 10^{-1}, 1$ for $N = 100000$ samples. Provide the curves of the autocorrelation function for each case in a single plot. Comment on the results, in particular the effect ϵ on mixing, provide a reasoning to explain it. (2 marks)

6. Now suppose that $p(y|\theta)$ cannot be computed in closed form. Derive a basic importance sampling (IS) estimator for this quantity. (1 mark)

7. Fix $\epsilon = 1$ for the rest of this question. Let us denote the IS estimator you derived in previous part with $\widehat{p(y|\theta)}$. We denote the relative absolute error (RAE) as

$$\mathsf{RAE}_{\sigma_{IS}, K} = \frac{\left|\widehat{p(y|\theta)} - p(y|\theta)\right|}{p(y|\theta)},$$

where σ_{IS} is the standard deviation of the proposal distribution $q = \mathcal{N}(0, \sigma_{IS}^2 I)$ and K is the number of samples used to build this IS estimator. To test this estimator, we will set $\theta = [0, 0]$ and use y as above. Compute the ground truth value $p(y|\theta)$, tune σ_{IS}^2, and demonstrate the estimator accuracy by computing RAE for $K = 10^2, 10^3, 10^4, 10^5$. To demonstrate this, you may need to run, e.g., $M = 10$ Monte Carlo simulations for each K and plot average relative absolute errors (over Monte Carlo simulations) w.r.t. K. Compare the resulting curve with a curve that decays like $1/\sqrt{K}$ visually. You do not need to get good relative errors here, but attempt at demonstrating the rate. Due to randomness, this curve may not exactly decay like $1/\sqrt{K}$, but it should be close enough. (4 marks)

8. Plug this IS estimator into the MH ratio to obtain a pseudo-marginal Metropolis Hastings (PMMH) method as shown in the lecture notes. Describe the method and the algorithm in detail. (2 marks)

9. For numerical results, you may have to tune σ_{IS}^2 and σ_{MH}^2 for the best performance (do not assume the values you optimized above will work). Set $N = 10000$ (set an appropriate `burnin` stage). Run the PMMH sampler for $K = 10^1, 10^2, 10^3$, provide autocorrelation functions for each case. Comment on the effect of K on mixing. Visualise the samples drawn for $K = 10^3$ case. How do these samples compare to the samples in Part 2? Elaborate. (5 marks)

Q2: IMPORTANCE SAMPLING FOR A HIDDEN MARKOV MODEL

Assume that we would like to understand the performance of a self-normalised importance sampler (SNIS) for the posterior sampling task $p(x_{0:n}|y_{1:n})$ in a state-space model.

Given a linear-Gaussian hidden Markov model:

$$p(x_0) = \mathcal{N}(x_0; m_0, P_0), \tag{1}$$
$$\tau(x_n|x_{n-1}) = \mathcal{N}(x_t; Ax_{n-1}, Q), \tag{2}$$
$$g(y_n|x_n) = \mathcal{N}(y_n; Hx_n, R), \tag{3}$$

for $n = 1, \ldots, T$. For fixed T, we assume the following scalar parameterization:

$$m_0 = 0, \quad P_0 = 1, \quad A = 0.9, \quad Q = 0.01, \quad H = 1, \quad R = 0.1.$$

1. Describe a procedure to simulate data from this hidden Markov model above and implement it. To demonstrate the implementation, simulate different lenghts of data $T = 2, \ldots, 20$ to demonstrate your function, plot $x_{0:T}$ vs. $y_{1:T}$ (observations, scattered) for each T. Note that each dataset is generated independently, i.e. the dataset with $T = 2$ is completely different from $T = 3$. When plotting, pay attention to indexing as states are indexed from 0 and observations are from 1. Store these datasets and do not modify them as they will be used below. (2 marks)

2. Describe a Kalman-based strategy to estimate the mean of the posterior distribution $p(x_{0:T}|y_{1:T})$, i.e.,

$$m_{0:T}^{(T)} = \mathbb{E}[x_{0:T}|y_{1:T}],$$

for each $T = 2, \ldots, 20$ using the datasets generated in Part 1. We denote this mean $m_{0:T}^{(T)}$ for $T = 2, \ldots, 20$ (we index the estimator with T to stress the fact that the datasets used for the estimators are different). (3 marks)

3. Implement this method and obtain the estimate $m_{0:T}^{(T)}$ for $T = 2, \ldots, 20$. Plot them vs. true states you have generated in Part 1 (this would be a separate plot for each $T = 2, \ldots, 20$). Store these estimates $m_{0:T}^{(T)}$, as they will be used below. (3 marks)

4. Derive a self-normalised importance sampling (SNIS) estimator for the quantity $m_{0:T}$, i.e., the mean of the posterior $p(x_{0:T}|y_{1:T})$. Use a proposal of the form

$$q(x_{0:T}) = \prod_{k=0}^{T} q(x_k),$$

where $q(x_k) = \mathcal{N}(0, \sigma_{IS}^2)$. Derive the SNIS estimator for the mean, denote it $\bar{m}_{0:T}^{(T)}$. (2 marks)

5. Run the estimator for $K = 10000$ particles for the dataset for $T = 2$ and record the SNIS estimate $\bar{m}_{0:2}^{(2)}$. Compare this to the estimate visually to the estimate you obtained in Part 2 for $T = 2$ namely $m_{0:2}^{(2)}$. You will have to tune σ_{IS}^2 for a good performance. (3 marks)

6. Run the SNIS estimator for $T = 2, \ldots, 20$. Choose an appropriate σ_{IS}^2 for a good performance (a good candidate is the one in previous part) and take $K = 10000$.

Provide two plots: (i) The effective sample size (ESS) value vs. $T = 2, \ldots, 20$ and (ii) the normalised mean squared error:

$$\mathsf{NMSE}(T) = \frac{\|\bar{m}_{0:T}^{(T)} - m_{0:T}^{(T)}\|^2}{\|m_{0:T}^{(T)}\|^2},$$

vs. $T = 2, \ldots, 20$. Comment on the results. What trend do you observe as T grows? What conclusions can you draw from this for estimating expectations with growing state-space dimension? (5 marks)