Exercise 12 – *Bayesian inference and Data assimilation*

Due by: Tuesday, 18 July 2023, 23:59 (CEST)

Problem 1 [Importance sampling] Consider a random variable $X \sim \pi$ that takes values in $\mathbb{R}^d$. We want to compute $\pi[g] := \mathbb{E}[g(X)]$ for some given function $g : \mathbb{R}^d \to \mathbb{R}$ by Monte Carlo method.

1. Suppose we have N iid samples of $X_i \sim \pi$. The sample mean is defined by

$$\bar{g}_X := \frac{1}{N} \sum_{i=1}^{N} g(X_i)$$

   Note that $\bar{g}$ is also a random variable. Find the variance of $\bar{g}_X$.

2. Consider another random variable $Y \sim \pi'$. Show that

$$\pi[g] = \mathbb{E}\Big[\frac{\pi(Y)}{\pi'(Y)} g(Y)\Big]$$

3. Use N iid samples of $Y_i \sim \pi'$ to compute the new sample mean

$$\bar{g}_Y := \frac{1}{N} \sum_{i=1}^{N} \frac{\pi(Y_i)}{\pi'(Y_i)} g(Y_i)$$

   Find the variance of $\bar{g}_Y$.

4. Assume $g(x) \leq 0$ implies $\pi(x) = 0$. In this case, find π' in terms of π and g that minimizes the variance of $\bar{g}_Y$. (Hint: try to make the expression a deterministic constant.)

   In practice, one would never achieve the optimal π'—remember the whole reason of doing this Monte Carlo method is to compute $\pi[g]$. In practice, you need to propose a suitable π', that potentially represent $\pi[g]$ well, or at least we know how to easily sample from.

5. [Importance resampling] Let us view step 3 in this way: First we sample $Y_i \sim \pi'$, and define weights by

$$\tilde{w}_i := C \frac{\pi(Y_i)}{\pi'(Y_i)}$$

   where C is any redundant constant, and we normalize $w_i := \dfrac{\tilde{w}_i}{\sum_i \tilde{w}_t}$. Now $\bar{g}_Y$ is the weighted average of $g(Y_i)$'s. Now we resample $\{X_i\}$ from $\{Y_i\}$ with probability mass function $\{w_i\}$.

   Suppose the proposal π' is Gaussian—or whatever we know how to sample from. Write a pseudo-code for the importance resampling to generate N samples from π. Assume You can generate samples from standard distributions: Gaussian, uniform, etc.

Problem 2 The Bayes inference can be viewed as a special case of important sampling, where π is the posterior and π' is our prior. Then the importance weight is proportional to the likelihood. We extend this to the discrete time filtering problem.
Let X be a Markov process given by a probability transition kernel

$$X_n \sim p_n(x_n \mid x_{n-1}), \quad n = 1, 2, \dots$$

and we are also given $X_0 \sim p_0$. The observation is given by

$$Y_n \sim q_n(y_n \mid x_n), \quad n = 1, 2, \dots$$

1. Write a pseudo-code for implementing particle filter that is explained below. Input is the number of particles N, time horizon T and the observation path $\{y_t : t = 1, \dots, T\}$. Output is $\{(x_T^i, w_T^i) : i = 1, \dots, N\}$ where the weighted empirical distribution approximates the posterior distribution $\pi_{X_T|Y_{1:T}}$.

   It should contain initialization, and iteration over time for the recursive step, and each recursive step may have iteration(s) over samples.

   - We start from N samples from the prior $X_0 \sim p_0$, say $\{x_0^i : i = 1, \dots, N\}$. The weight is given by [BLAH] for all i.
   - Given samples and weights $\{(x_{n-1}^i, w_{n-1}^i) : i = 1, \dots, N\}$, the current approximation of the posterior is given by:

     $$\pi_{X_{n-1}|Y_{1:n-1}} \approx \sum_{i=1}^{N} w_{n-1}^i \delta(x_{n-1}^i)$$

     where $\delta(x)$ denotes the Dirac delta distribution at x. We draw new samples from $x_n^i \sim p_n(x_n \mid x_{n-1}^i)$; and then we may claim

     $$\pi_{X_n|Y_{1:n-1}} \approx \sum_{i=1}^{N} w_{n-1}^i \delta(x_n^i)$$

   - With the observation $Y_n = y_n$, find the new weight in terms of w_{n-1}^i, x_n^i and y_n. You can think of this as the importance sampling in the case where π' is $\pi_{X_n|Y_{1:n-1}}$ and π is $\pi_{X_n|Y_{1:n}}$.

2. We have already discussed about weight degeneracy for particle filtering, briefly in P8 of Ex7 and qualitatively in P2 of Ex11. Indeed, resampling is the key step to maintain the effective sample size large enough. Add a resampling step to your pseudo-code.

   - We need one more input $\gamma \in (0, 1)$.
   - Check if ESS:$= \frac{1}{\sum_i (w_n^i)^2}$ is less than γN.
   - How to obtain new samples from current weights (more or less step 5 of P1).
   - What is the new value for the weights?

Problem 3 Implement the importance resampling algorithm that we discussed in P1. Suppose you would like to sample from the following mixture of Gaussian random variables:

$$\pi(x) = \frac{1}{2}N(-4, 1) + \frac{1}{2}N(4, 1)$$

1. Generate 10,000 samples of X using Gaussian proposal $\pi' = N(0, \sigma^2)$ for both $\sigma^2 = 1$ and $\sigma^2 = 4$.

2. Plot histograms of these samples, and overlay with the true density π. Discuss the consequences of each choice of σ^2.