**Exercise 8** – *Bayesian inference and Data assimilation*

**Due by:** Tuesday, 13 June 2023, 23:59 (CEST)

**Problem 1** In this exercise, you will implement the Euler-Maruyama method. It is very similar to the Euler scheme that you implemented in the exercise 1, but this time we are interested in stochastic differential equations (SDEs) instead of ODEs.
Consider a continuous time process $\{X_t \in \mathbb{R} : t \geq 0\}$ governed by the SDE:

$$\mathrm{d}X_t = -X_t \, \mathrm{d}t + \sqrt{2} \, \mathrm{d}B_t, \quad X_0 = x_0$$

Or in its integral form,

$$X_t = x_0 - \int_0^t X_s \, \mathrm{d}s + \sqrt{2} \, B_t$$

where $\{B_t : t \geq 0\}$ is a standard Brownian motion. In order to apply Euler scheme, consider a time-discretization $\Delta t$ and approximate

$$X_{t+\Delta t} - X_t \approx -X_t \Delta t + \sqrt{2}\big(B_{t+\Delta t} - B_t\big)$$

The idea is to exploit the fact that the distribution of $B_{t+\Delta t} - B_t$ is Gaussian with mean 0 and variance $\Delta t$. Hence we discretize the SDE by

$$x_{n+1} = x_n - x_n \Delta t + \sqrt{2\Delta t} \, \xi_n$$

where $\xi_n \sim N(0, 1)$ for $n = 0, 1, \ldots$.

1. What is the distribution of $x_1$ and $x_2$?

2. Set $\Delta t = 0.01$ and $x_0 = 2$. Run 10,000 Monte-Carlo simulations of above. Save and plot histogram of the empirical distributions at time $t = 0$, $t = 0.5$, $t = 1$ and $t = 10$.

3. How do the distribution at the different times look? Do you see if it converges to something? You may run the simulation further in time if necessary.

4. ~[Digression] This process is an example of so-called Ornstein-Uhlenbeck process. It converts any (complicated) distribution $\mu$ of $X_0$ into a certain distribution you would find in step 3. Suppose you can run this process reverse-in-time, and then what you can do with this?

**Problem 2** We now use the above method to sample different distribution. Assume you are measuring a size of members of some animal species. The distribution of the height is given as

$$\pi(x) = \frac{1}{C} \exp\big(-V(x)\big)$$

with

$$V(x) = \big((x-4)^2 - 2\big)^2$$

and $C$ is the normalization constant such that $\int_{\mathbb{R}} \pi(x) \, \mathrm{d}x = 1$.

1. Plot $\pi$ on suitable interval with $C = 1$.
2. Consider the following SDE:

$$\mathrm{d}X_t = -\nabla V(X_t)\,\mathrm{d}t + \sqrt{2}\,\mathrm{d}B_t, \quad x_0 = 1$$

   Run 10,000 Monte-Carlo simulations with $\Delta t = 0.01$ up to time $T = 100$.
3. Compare the histogram of simulated samples and $\pi(x)$ by using suitable constant factor.
4. Using the samples, what do you expect to be the amount of animal who have a size greater than 6?

**Problem 3** Consider a discrete-time Markov process whose pdf evolves according to

$$\pi_{n+1}(x') = \int_{\mathbb{R}} P(x;x')\pi_n(x)\,\mathrm{d}x$$

We are interested in the distribution from two different prior $\pi_0$ and $\tilde{\pi}_0$. Denote the distribution starting from $\tilde{\pi}_0$ by $\tilde{\pi}_n$.

1. Show that

$$d_{\mathrm{TV}}(\pi_1, \tilde{\pi}_1) \le d_{\mathrm{TV}}(\pi_0, \tilde{\pi}_0)$$

   where the total variation distance $d_{\mathrm{TV}}(\mu,\nu)$ defined by

$$d_{\mathrm{TV}}(\mu,\nu) = \frac{1}{2}\sup_{|f|_\infty \le 1} \left|\mathbb{E}_\mu[f] - \mathbb{E}_\nu[f]\right|$$

2. Conclude that the distance between the trajectory from two different prior is uniformly bounded, that is,

$$d_{\mathrm{TV}}(\pi_n, \tilde{\pi}_n) \le d_{\mathrm{TV}}(\pi_0, \tilde{\pi}_0), \quad \forall\, n = 0, 1, \ldots$$