Bayesian inference and data assimilation, SS2019
Dr. Ester Mariucci, Dr. Han Cheng Lie, Mr. Jens Fischer
Institut für Mathematik, Universität Potsdam

If you have questions about the exercise sheet, write to hanlie@uni-potsdam.de. E-mails will be replied to during working hours.

Exercise Sheet 1 Due date: Thursday, 25.04.2019

Exercise 1. Using the file 'Example-1-1.m' from the Moodle site, implement the numerical model from the Lorenz-63 example with step-size $\delta t = 0.01$, starting from the initial condition $z(0) := (x_0, y_0, \tilde{z}_0)^\top = (-0.587, -0.563, 16.870)^\top$. Store the resulting reference trajectory $z(\mathsf{t}_k)$ using $\Delta t_{\text{out}} = 0.05$ between $t = 0$ and $t = 200$ in a file for Exercise 2 and 3 below in a matrix of size 3×4001. Plot the projection of the data set to the (x, y) plane. Submit your code. Note: the comments already included in the file 'Example-1-1.m' give examples of how to write comments.

(Plot - **0.5 point**; Code performs task - **1 point**; Comments - **0.5 point**; Total 2 points)

Exercise 2. Using the reference trajectory of $z(t)$ that you generated in Exercise 1 and using the forward map $h((x, y, \tilde{z})^\top) = x$ (i.e. projection to the first coordinate), generate the observation process

$$y_{\text{obs}}(\mathsf{t}_k) = \mathrm{x}_{\text{obs}}(\mathsf{t}_k) + \frac{1}{20} \sum_{i=1+(k-1)20}^{20k} \eta_i, \quad k = 1, \ldots, N_{\text{obs}} = 4000$$

where $\{\eta_i\}_{i \in \mathbb{N}_0}$ is a sequence of real numbers defined by tent map iteration, with initial state $\eta_0 = 4(2^{-1/2} - \frac{1}{2})$ and

$$\eta_{i+1} = \begin{cases} 1.9999\eta_i + \frac{a}{2} & -\frac{a}{2} \le \eta_i < 0 \\ -1.9999\eta_i + \frac{a}{2} & 0 \le \eta_i \le \frac{a}{2} \end{cases}$$

with $a = 4$ in a file for later use. Submit your code.

(Code performs task - **0.5 point**; Comments - **0.5 point**; Total 1 point)

Exercise 3. Produce forecasts for the forecast interval $\Delta t_{\text{out}} = 0.05$ using the observations produced in Exercise 2 and the linear interpolation polynomial

$$q(t) = y_{\text{obs}}(\mathsf{t}_k) + (t - \mathsf{t}_k) \frac{y_{\text{obs}}(\mathsf{t}_k) - y_{\text{obs}}(\mathsf{t}_{k-1})}{\mathsf{t}_k - \mathsf{t}_{k-1}}$$

to obtain your forecast $y_{\text{pred}}(\mathsf{t}_{k+1})$ for $k = 1, \ldots, N_{\text{obs}}$. Compute the time averaged RMSE. Plot your observations and your predictions in one figure. Repeat this procedure for the forecast interval $2\Delta t_{\text{out}} = 0.1$. State your conclusion from these experiments and submit your code.

(Plots - **1 point**; Code performs task - **1 point**; Comments - **0.5 point**; Conclusion - **0.5 point**; Total 3 points)

Exercise 4. Let X, Y, Z be $\mathbb{R}$-valued random variables on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ with finite second moments, i.e. $\mathbb{E}[X^2]$, $\mathbb{E}[Y^2]$, and $\mathbb{E}[Z^2]$ are finite. Let $\alpha, \beta \in \mathbb{R}$ be fixed, deterministic (i.e. non-random) numbers. Recall that the **variance** of a random variable with finite second moment is defined by

$$\mathbb{V}[X] := \mathbb{E}[(X - \mathbb{E}[X])^2]$$

and that the **covariance** of two random variables is defined by

$$\mathbb{C}[X, Y] := \mathbb{E}[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])].$$

(a) Complete the following equation: $\mathbb{E}[\alpha X + \beta] =?$ (**0.5 point**)

(b) Use your answer from (a) to prove that $\mathbb{C}[X, Y] = \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y]$. Use this to prove that $\mathbb{V}[X] = \mathbb{E}[X^2] - \mathbb{E}[X]^2$. (**0.5 point** for correct proofs of both equations only; otherwise 0 points)

Exercise 5. Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space, and let $X : \Omega \to \mathcal{X}$ and $Y : \Omega \to \mathcal{Y}$ be $\mathbb{R}$-valued discrete random variables such that $\mathcal{X}$ and $\mathcal{Y}$ are finite subsets of $\mathbb{R}$. Two discrete random variables X and Y with values in $\mathcal{X}$ and $\mathcal{Y}$ respectively are said to be **independent** if for any $x \in \mathcal{X}$ and $y \in \mathcal{Y}$, the following equation is satisfied:

$$\mathbb{P}(X = x, Y = y) = \mathbb{P}(X = x)\mathbb{P}(Y = y). \tag{1}$$

In words, we say that the joint probability (the quantity on the left-hand side) **factorises** into the marginal probabilities (the quantities on the right-hand side).

Suppose that $\mathcal{X} = \{1, 2, 3\}$ and $\mathcal{Y} = \{2, 5, 7\}$, and you are given the following table of joint probabilities.

		X		
		x=1	x=2	x=3
Y	y=2	0.05	0.03	0.02
	y=5	0.2	0.12	0.1
	y=7	0.25	0.15	0.08

Table 1: Joint probabilities of the discrete random variables X and Y.

(a) Calculate the marginal probabilities of X and of Y, i.e. compute $\mathbb{P}(X = x)$ for $x = 1, 2, 3$ and $\mathbb{P}(Y = y)$ for $y = 2, 5, 7$. (**1.5 points**)

(b) Prove either that X and Y are independent by verifying that (1) is satisfied for all possible combinations of x and y, or that X and Y are not independent by finding x and y such that (1) is not satisfied. (**1.5 points**)