3. EXERCISE SHEET *Bayesian inference and data assimilation*

Exercise 1. On Sheet 2 Exercise 3, you used linear extrapolation to predict the next instance of the Lorentz system, however, we did not actually use any knowledge about the model that generated the data. Now we take a different approach and incorporate model information into our prediction. This is a first step into the direction of data assimilation. Use $\delta t = 0.01$ to simulate the trajectory and save the output on every fifth step, i.e. $\Delta t = 0.05$.

(i) Compute a Lorenz-63 trajectory starting at the same initial conditions as on sheet 2 but do not include the noise terms from the tent map. Compute the RMSE when comparing this trajectory to the true trajectory. Calculate one RMSE for all three coordinates, not one RMSE for each coordinate. Do not forget to divide the RMSE by 3 to make it comparable to the RMSE for only the x-coordinate on sheet 2. Overlay the true and the approximated trajectory in one plot. Discuss your findings.

(ii) Use the observations you generated in Exercise 2 on the last sheet. Do the same as in part (i) but on every fifth steps include the observations into the current state by setting the x-coordinate of your system to the observed x-coordinate. Plot both of the plots from part (i) and the newly generated trajectory from this exercise into one plot. Compute the RMSE again. Discuss your findings.

When saying plot the trajectory in this exercise please do three plots - one for each coordinate x, y, z. Do not plot a 3-dimensional plot with all coordinates.

Exercise 2. Consider two univariate Gaussian random variables X and Ξ and define a third random variable Y for given coefficient g as follows:

$$Y = gX + \xi.$$

Provided that $g \neq 0$, then $X = (Y - \Xi)/g$. Instead of this simple rearrangement, we attempt to approximate X in terms of Y alone and make the linear *ansatz*

$$\hat{X} = cY + d.$$

Determine the unknown coefficients c and d through minimisation of the expected distance between X and $\hat{X}$, *i.e.* $\mathbb{E}[(X - \hat{X})^2]$. Assume that $X \sim \mathrm{N}(\mu, \sigma^2)$ and $\xi \sim \mathrm{N}(0, r^2)$ are independent.

Exercise 3. Let X_1 and X_2 be two random variables with joint PDF

$$\pi_{X_1 X_2}(x_1, x_2) = \frac{1}{Z} \exp\left(-x_1^2 - x_2^2 - x_1^2 x_2^2\right)$$

where Z is a normalisation constant. Evaluate $\mathbb{E}[X_1 X_2^2 | X_1 = a]$.

Exercise 4. Consider the two-dimensional Gaussian PDF $\mathrm{n}(z; \bar{z}, P)$, $z = (x_1, x_2)$ with mean $\bar{z} = (\bar{x}_1, \bar{x}_2)$ and covariance matrix

$$P = \begin{pmatrix} \sigma_{11}^2 & \sigma_{12}^2 \\ \sigma_{21}^2 & \sigma_{22}^2 \end{pmatrix}.$$

Verify that

$$\begin{aligned} \mathrm{n}(z; \bar{z}, P) &= \frac{1}{2\pi |P|^{1/2}} \exp\left(-\frac{1}{2}(z - \bar{z})^T P^{-1} (z - \bar{z}) \right) \\ &= \frac{1}{\sqrt{2\pi}\sigma_c} \exp\left(-\frac{1}{2\sigma_c^2}(x_1 - \bar{x}_c)^2 \right) \frac{1}{\sqrt{2\pi}\sigma_{22}} \exp\left(-\frac{1}{2\sigma_{22}^2}(x_2 - \bar{x}_2)^2 \right). \end{aligned}$$

What are the corresponding formulas for the conditional PDF $\pi_{X_2}(x_2|x_1)$ and the marginal $\pi_{X_1}(x_1)$?

Exercise 5. Show that the Hellinger distance

$$d_{\mathrm{Hell}}(p, q) = \left(\frac{1}{2} \int \left(\sqrt{p(x)} - \sqrt{q(x)} \right)^2 \mathrm{d}x \right)^{1/2}$$

and the Kullback-Leibler divergence

$$D_{\mathrm{KL}}(p||q) = -\int \log \frac{q(x)}{p(x)} \, p(x) \mathrm{d}x$$

satisfy the inequality

$$d_{\mathrm{Hell}}(p, q)^2 \leq \frac{1}{2} D_{\mathrm{KL}}(p||q).$$

You may use that $2(1 - \sqrt{x}) \leq -\log x$.