Data Assimilation (summer term 2025)

Exercise sheet

June 16 + June 23, 2025

Bayesian Inference, MCMC

1. Task.
We solve a toy Bayesian inference problem. Consider the hypothesis random variable H with prior $H \sim \mathrm{N}(1, 1)$. The observable is given by evidence E:

$$E = H^2 + W$$

where $W \sim \mathrm{N}(0, 1)$. The random variables H and W are assumed to be independent.

a) Using Bayes' formula, find the expression for $\pi_{H|E=e}(h)$, the conditional density of H at h given $E = e$. You may omit the explicit formula for the normalization constant.

b) Plot the conditional PDF given $e = 2$ and find the *maximum a posteriori* (MAP) estimator of H given $E = e$.

2. Task.
Suppose you are trying to estimate the true temperature at a location, denoted by $x \in \mathbb{R}$. Its estimate is a random variable denoted by X. You have two sensors that provide noisy measurements of the temperature:

- Sensor 1 reports twice the true temperature plus noise.
- Sensor 2 reports the true temperature plus noise.

Mathematically, the measurements are modeled as:

$$Y = \begin{pmatrix} Y_1 \\ Y_2 \end{pmatrix} = \begin{pmatrix} 2 \\ 1 \end{pmatrix} X + W,$$

where $W \sim \mathcal{N}(0, \gamma^2 I_2)$ represents independent Gaussian noise in both sensors, and I_2 is the 2×2 identity matrix.

Before collecting any data, you believe the temperature is likely close to zero, but could reasonably vary. You model this prior belief as $X \sim \mathcal{N}(0, 2)$.

Suppose you observe the following sensor readings: $y = \begin{pmatrix} 1 \\ 2 \end{pmatrix}$.

a) Compute the posterior distribution for the true temperature x given these sensor readings.

b) What happens to the posterior distribution and variance as the sensor noise γ decreases (i.e., as the sensors become more accurate)?

c) How do you interpret the limiting distribution in this context?

3. Task.
A public health researcher wants to estimate the average systolic blood pressure (SBP) of adults in Freiberg. However, only 20 adults took part in the study. To estimate the true average SBP μ in Freiberg and to quantify the uncertainty she wants to use Bayesian statistics.

- **Observation**: These SBP measurements of the 20 adults from Freiberg are $x_1, x_2, \ldots, x_{20}$.
- **What we know**: Based on medical literature, the standard deviation of SBP in adults is about $\sigma = 15$ mmHg [1]. From previous national studies she knows that the average SBP is about $\mu_{\text{Germany}} = 130$ mmHg.

[1] Please note that the figures are fictitious and only part of the example.

She defines a prior (initial belief about the estimate) and a likelihood for the **model**.

- **Prior** for mean SBP μ in Freiberg: We believe, before collecting data, that μ is likely close to μ_{Germany}, so we start from that and model the known uncertainty about it by:

$$\mu \sim \mathcal{N}(130, 15^2)$$

- **Likelihood**: Each measured SBP is assumed to be normally distributed around μ, with known standard deviation: $x_i \mid \mu \sim \mathcal{N}(\mu, 15^2)$
- **Evidence**: $\pi(x_1, x_2, \ldots, x_{20})$ is the marginal likelihood

a) How is the posterior $\pi(\mu \mid x_1, x_2, \ldots, x_{20})$ generally defined? What is the resulting posterior distribution in our case?

b) Another measurement x_{21} needs to be included. Based on $\pi(x_1, x_2, \ldots, x_{20})$ how is the new posterior $\pi(\mu \mid x_1, x_2, \ldots, x_{21})$ generally defined? What is the resulting posterior distribution?

c) How does the uncertainty about μ change with new data points?

d) Assume the (unknown) true SBP mean of Freiberg was $\mu = 125$ mmHg. Please implement the **Metropolis–Hastings algorithm** in Python to approximate the posterior $\pi(\mu \mid x_1, x_2, \ldots, x_{20})$ using the given model for the posterior.

   For the evidence generate a synthetical data set of $x_i \sim \mathcal{N}(125, 15^2)$, $i = 1, \ldots, 20$.

   Compare the results with the analytical solution from 3a).