
Problem for Day 3. *Consider the statistical inverse problem*

$$Y = \mathcal{A}X + E, \tag{1}$$

where $X \in \mathbb{R}^2$ *and* $Y \in \mathbb{R}^3$ *are real-valued multivariate random variables and* $\mathcal{A} \in \mathbb{R}^{3\times 2}$ *is a matrix given by*

$$\mathcal{A} = \begin{pmatrix} 1 & -1 \\ 1 & -2 \\ 2 & 1 \end{pmatrix}.$$

Further more, suppose that the prior X *and noise* E *follow the distributions* $X \sim \mathcal{N}(0, I_2)$ *and* $E \sim \mathcal{N}(0, \sigma^2 I_3)$*, where* I_2 *and* I_3 *are the 2 and 3 dimensional identity matrix, respectively.*

1. *Use Bayes' theorem to write the density function of the posterior distribution* $\pi_{X|Y=\mathbf{y}}(\mathbf{x})$ *up to a proportionality constant:*

$$\pi_{X|Y=\mathbf{y}} \propto \cdots$$

   only express it in terms of $\mathcal{A}$, $\mathbf{x}$, $\mathbf{y}$, σ.

2. *open the code* `exercise.py` *in day 3 folder and follow the following instructions*
   (a) *write a function* `prior` *that computes the prior probability density of an input* `x`*.*
   (b) *define a Numpy matrix* `A` *that holds* $\mathcal{A}$*.*
   (c) *write a functions* `likelihood` *and* `posterior` *that computes the likelihood probability density and posterior probability density for an input* `x`*. Here use* `sigma` *and* `y_obs` *that is provided in the code.*
   (d) *Complete the code for the Metropolis-Hastings random walk algorithm based on transition strategy*

$$\mathbf{x}^\star = \mathcal{N}(\mathbf{x}, c^2 I_2),$$

   or equivalently

$$\mathbf{x}^\star = \mathbf{x} + c\mathbf{z}.$$

   Here, $\mathbf{z}$ *is a sample from* $\mathcal{N}(0, I_2)$ *and* c *is the step-size in the random walk Metropolis-Hastings algorithm.*

3. *Draw 50000 samples from the posterior distribution. Down-sample (skip every 10 samples) to create a near i.i.d. samples of the posterior. plot a 2D histogram of the posterior and mark the posterior mean on it.*

4. *Let the step size be* $c \in \{0.001, 0.1, 10\}$*. Plot the posterior 2D histogram for each step size and explain the differences. In your opinion, which value of* c *is better for this problem, and why?*

5. *Let noise variance be* $\sigma^2 \in \{0.01, 0.1, 1\}$*. Plot the posterior 2D histogram for each noise variance. What differences do you observe? Discuss uncertainty in the posterior mean estimation.*