Inverse Problems — Exercise 10

Sommersemester 2023

Return your written solutions either in person or by email to vesa.kaarnioja@fu-berlin.de by Tuesday 11 July, 2023, 10:15

Please note that there are a total of 4 tasks in this exercise sheet.

1. Assume that you have a Gaussian posterior distribution

$$\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \sim \pi^y \sim \mathcal{N}\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & p \\ p & 1 \end{pmatrix}\right).$$

   (a) Write a Gibbs sampler for the posterior π^y. Based on the generated samples, what are the conditional mean estimate of π^y and the marginal standard deviations of x_1 and x_2?

   (b) Repeat part (a) for parameter values $p = 0.5, 0.9, 0.99$, and 0.999. How does the degree of correlation between x_1 and x_2 affect the performance of the Gibbs sampler?

2. Suppose we have an inverse problem

$$y = \begin{pmatrix} x_1^2 + x_2^2 \\ x_2 \end{pmatrix} + \eta,$$

   where $y \in \mathbb{R}^2$, $x = (x_1, x_2)^{\mathrm{T}} \in \mathbb{R}^2$. Let us set the prior $x = z \cdot \mathbf{1}_{[-4,4]^2}(z)$, where $z \sim \mathcal{N}((0,0)^{\mathrm{T}}, I)$,

$$\mathbf{1}_B(z) = \begin{cases} 1, & z \in B, \\ 0, & \text{otherwise,} \end{cases}$$

   and $\eta \sim \mathcal{N}(0, \delta^2 I)$ with $\delta = 0.1$. Suppose we are given the observation $y = (7, -2)^{\mathrm{T}}$. Implement MCMC with Metropolis–Hastings kernel

$$x_{k+1} \sim \sqrt{1-\beta^2} \cdot x_k + \beta\xi, \quad \xi \sim \mathcal{N}(0, I),$$

   for different values of $\beta \in (0,1)$ to sample the posterior density. For each value of β produce 10 000 samples and plot them. What do you notice? Also compute for each β the *acceptance ratio*, i.e., the ratio between accepted jumps and the total length of the chain. Use the origin as initial value.

   Using the best choice of β, compute the expectation of the posterior, i.e., the conditional mean estimate

$$x_{\mathrm{CM}} = \int_{\mathbb{R}^2} x\pi^y(x)\,\mathrm{d}x.$$

3. Let $\rho_1 : \mathbb{R} \to \mathbb{R}_{>0}$ and $\rho_2 : \mathbb{R} \to \mathbb{R}_{>0}$ be probability densities. The Kullback–Leibler divergence is defined by

$$d_{\mathrm{KL}}(\rho_1 \| \rho_2) = \int_{\mathbb{R}} \log\left(\frac{\rho_1(x)}{\rho_2(x)}\right) \rho_1(x)\,\mathrm{d}x. \tag{1}$$

   Solve the Kullback–Leibler divergence $d_{\mathrm{KL}}(\rho_1 \| \rho_2)$ for $\rho_1 \sim \mathcal{N}(m_1, \sigma_1^2)$ and $\rho_2 \sim \mathcal{N}(m_2, \sigma_2^2)$, where $m_1, m_2 \in \mathbb{R}$ and $\sigma_1, \sigma_2 > 0$.

4. Let $\rho \sim \mathcal{N}(0,1)$ and assume that the posterior density is given by

$$\pi^y(x) = 0.6 \cdot \rho(x-3) + 0.4 \cdot \rho(x+2).$$

Using your favorite programming language, try to find the KL-optimal approximation for π^y in the class of $\mathcal{B} = \{\mathcal{N}(m,1) \mid m \in \mathbb{R}\}$, that is, find m that solves

(a) $\inf_{m\in\mathbb{R}} d_{\mathrm{KL}}(\mathcal{N}(m,1)\|\pi^y)$,

(b) $\inf_{m\in\mathbb{R}} d_{\mathrm{KL}}(\pi^y\|\mathcal{N}(m,1))$,

where d_{KL} is the Kullback–Leibler divergence defined in (1). In each case, plot the probability density for $\mathcal{N}(m,1)$ against the posterior density π^y. One of the approximations you obtained should be centered around the mode of π^y while the other one should be centered around the mean of π^y. Which one is which?

Hint: You are not required to do anything too sophisticated in this task. Any reasonable minimization procedure or quadrature rule is OK.