CWI Autumn School 2024: Numerical Methods for Bayesian Inverse Problems

Exercise Sheet 2

Exercise 2: Numerical Stability of the Posterior.

We consider the logistic differential equation

$$\dot{x}(t) = r\, x(t)\,(k - x(t)), \qquad x(0) = x_0\,,$$

describing the growth of a population in a habitat with capactity k. The initial state $x_0 = 0.1$ and the reproduction rate $r = 0.25$ are known. However, we are uncertain about the capacity k.

We propose a uniform prior for $k \sim \pi_0 = \mathsf{U}[10, 20]$ and aim to update that prior given noisy observations $y = (y_1, y_2, y_3)^\top \in \mathbb{R}^3$ of

$$y_i = x(i) + \eta_i, \qquad \eta_i \sim \mathsf{N}(0, 1)$$

at the 'times' $i = 1, 2, 3$ with stochastically independent η_i. The given data is $y = (3, 14, 17)^\top$.

The logistic equation can be solved analytically

$$x(t) = \frac{k}{1 + \exp(-r\,k\,t)\left(\frac{k}{x_0} - 1\right)}, \qquad t \geq 0,$$

but to explore the stability of the posterior distribution with respect to approximations of the forward operator, we will solve it numerically using a time stepping scheme.

Euler's method: Consider

$$\dot{x}(t) = f(t, x(t)) \qquad x(0) = x_0\,.$$

Then, choose a stepsize $h > 0$ and generate approximations $x_i \approx x(t_i)$ for $t_i = i\,h$, $i \in \mathbb{N}_0$, recursively by

$$x_{i+1} = x_i + h\, f(t_i, x(t_i)).$$

Under suitable assumptions, the Euler method has a linear convergence rate, i.e.,

$$\sup_{i=1,\dots,T/h} |x_i x(t_i)| \propto h.$$

Thus, using the analytic solution we get the true posterior $\pi_{k|y}$ and using the Euler method we get approximate posteriors $\pi^h_{k|y}$ on the uncertain carrying capacity k.

(a) Using Bayes' rule, explicitly compute and plot the (normalized) posterior (densities) $\pi_{k|y}$ and $\pi^h_{k|y}$ for

$$h \in \{0.5, 0.25, 0.1, 0.05, 0.025, 0.01, 0.005, 0.0025, 0.001\}.$$

Comments: You may use the Matlab files `Euler.m` and `Exercise_01_template.m` provided in the git repository. To estimate the normalizing constant (i.e., the reciprocal of the evidence), you may use numerical quadrature on $[10, 20]$, e.g., the trapezoidal rule

$$\int_a^b f(k)\mathsf{d}k \approx \frac{\Delta k}{2} f(a) + (\Delta k)\sum_{j=1}^{N-1} f(a + j\,\Delta k) + \frac{\Delta k}{2} f(b), \qquad N = \frac{b-a}{\Delta k}$$

for, e.g., $\Delta k = 0.01$.

(b) Compute the corresponding Hellinger distances $d_{\mathsf{Hell}}(\pi_{k|y}, \pi^h_{k|y})$ and plot them in a log-log-plot versus the stepsize h in order to verify

$$d_{\mathsf{Hell}}(\pi_{k|y}, \pi^h_{k|y}) \propto h$$

in accordance to the theoretical result from the lectures.

Exercise 3: MCMC Performance.

We consider again the logistic differential equation

$$\dot{x}(t) = r\,x(t)\,(k(t) - x(t)), \qquad x(0) = x_0\,,$$

describing the growth of a population in a habitat, but now with time-dependent capactity $k(t)$, modelling the influence of the seasons on resources in the habitat. The initial state is now $x_0 = 1$ and the reproduction rate is again $r = 0.25$.

We want to infer the capacity k again. To this end we observe the data $y_i = x(i) + \eta_i$ with stochastically independent $\eta_i \sim \mathsf{N}(0, 0.25)$ at 'times' $i = 1, \ldots, 12$.

For your experiments you may use the Matlab template `Exercise_03_template.m` provided in the git repository. The logisitic equation is solved numerically using Euler's method on $[0, 12]$ with stepsize $h = 0.01$ using the special function `EulerLogEq.m` also provided. This function just needs to know $k(t_i)$ at times $t_i = i\,h$, $i = 0, \ldots, 1200$, i.e., it requires a vector $\boldsymbol{k} \in \mathbb{R}^{1201}$ as input.

(a) In order to generate (synthetic data) and for comparison we take as ground truth

$$k_{\mathsf{true}}(t) = 12 \exp\left(-\frac{1}{4}\cos\left(\frac{\pi}{6}t\right)\right).$$

In our simulations we add noise to the true observations and set the seed `rng(17)` for the pseudo random number generator in Matlab when drawing $\eta_1, \ldots, \eta_{12}$.

Use Euler's method to compute the ground truth $x(t)$ in $t \in [0, 12]$ and then plot it together with the perturbed data $y_1, \ldots, y_{12}$.

(b) We use a *lognormal* pior for k given by the series expansion

$$\log(K(t)) = \log(10) + \sum_{j=0}^{\infty} U_j\ \phi_j(t), \qquad U_j \sim \mathsf{N}(0, \lambda_j)$$

where $\phi_0(t) \equiv 1/12$, $\lambda_0 = \frac{1}{\pi^2}$, and $\phi_j(t) = \frac{1}{\sqrt{6}} \sin\big(j \frac{\pi}{12} t\big)$, $\lambda_j = \frac{1}{\pi^2 j^2}$, for $j \geq 1$.

In our simulations, we truncate the expansion after $N = 1000$ terms. Further, we need to know the ϕ_j only at the grid points $t_i = i\,h$, $i = 0, \ldots, 1200$. Thus, our Bayesian inference problem is to infer the coefficients $U = (U_1, \ldots, U_N)^\top$ wher the prior distribution is

$$U \sim \pi_0 = \mathsf{N}(0, \Lambda), \qquad \Lambda = \mathrm{diag}(\lambda_1, \ldots, \lambda_N)$$

Plot 100 prior realizations for k and compare them visually to the ground truth.

(c) Run an MCMC simulation using the pCN or random walk proposals discussed in lectures and use it to compute the posterior mean of the following quantity of interest

$$f(k) = \int_0^{12} k(t)\mathsf{d}t\,.$$

Tune the stepsize to an average acceptance rate of roughly 0.25 before running the full Markov chain simulation with $20{,}000$ iterations. **Comments:** You can again apply the trapezoidal rule to approximate $f(k)$. To estimate the average acceptance rate, you can simply average the values of the acceptance probability $\alpha(u_i, v_i)$ over the iterations.

(d) Store the states of the Markov chain in a matrix and use them to plot (the last) 100 realizations of the posterior for k. Compare them visually to the ground truth. Also plot the pointwise mean and the (centered) 95% credibility intervalls.

(e*) Compute an asymptotic confidence interval for $\mathbb{E}_{\pi_{k|y}}[f]$ using `autocorr` to estimate the autocorrelation.