Math. Meth. for UQ in Hydrology
Block 1
Exercise

Identification of the parameters of a rating curve

1 Objectives

In this exercise we will compute the parameters for a rating curve. For doing this we will apply inverse Bayesian regression and MCMC. In addition, we will analyze the impact of the choice of the parameters on the performance of the Random Walk Metropolis Algorithm (RWMA).

2 Formal setting

The data we are going to analyze are real data collected at the Plima stream between July and October 2019.

We have 11 measurements of water high (cm), indicated by the vector $\mathbf{h}$, and flow (l/s), indicated by the vector $\mathbf{q}$. Notice that in this script we decided to keep indicating flow values with the letter $\mathbf{q}$, as used in hydrology. This is not linked to the proposal distribution $\mathbf{Q}$ in the first script about MCMC, which will be then indicated with $\mathbf{S}$.

We assume that the rating curve has an exponential form. In order to deal with a linear problem we apply the natural logarithm and obtain

$$\log(\mathbf{q}) = a + b\log(\mathbf{h}) + \eta \tag{1}$$

where $\eta \sim \mathcal{N}(0, \Gamma)$ with $\Gamma \in \mathbb{R}^{n_d \times n_d}$. This can be rewritten as

$$\log \mathbf{q} = A\mathbf{x} + \eta, \tag{2}$$

where $A \in \mathbb{R}^{n_d \times n}$, with $A_{i1} = 1$ for $i = 1, ..., n_d$, $A_{i2} = \log(\mathbf{h}_i)$ and $\mathbf{x} = (a, b)$ is the parameter array (hence in this case $n = 2$). In our case study we have 11 measured data, so $n_d = 11$.

Prior

We assume to have a Gaussian prior, in particular

$$\rho_0(\mathbf{x}) \propto \mathcal{N}(\mathbf{m}_0, \Sigma_0), \tag{3}$$

where $\mathbf{m}_0 = [0, 0]$ and $\Sigma_0 = \begin{pmatrix} 10 & 0 \\ 0 & 10 \end{pmatrix}$.

Noise

We assume to have a Gaussian observational noise with 0 mean and covariance matrix Γ. We fix a 3% noise level and we define Γ as a diagonal matrix with $\Gamma_{ii} = 0.03 * \log(\mathbf{q}_i)$ for $i = 1, ..., n_d$, where $\log(\mathbf{q}_i)$ is the i-th component of the data vector $\log(\mathbf{q})$.

Likelihood

Since we assumed that the observational noise η is Gaussian distributed with 0 mean and Γ covariance matrix, we can write the likelihood as

$$\rho_{D|X}(\log(\mathbf{q})|\mathbf{x}) \propto \exp\left(-\frac{1}{2}|| \log(\mathbf{q}) - A\mathbf{x}||_\Gamma^2\right). \tag{4}$$

Posterior distribution

Thanks to the Bayes' Theorem, we have that the posterior is proportional to the product of the likelihood and the prior

$$\begin{aligned}\rho_{X|D}(\mathbf{x}|\log(\mathbf{q})) &\propto \rho_{D|X}(\log(\mathbf{q})|\mathbf{x})\rho_0(\mathbf{x}) \\ &\propto \exp\left(-\frac{1}{2}|| \log(\mathbf{q}) - A\mathbf{x}||_\Gamma^2\right)\rho_0(\mathbf{x}) \end{aligned}\tag{5}$$

In the particular case we are considering, we can compute the analytical form of the posterior and we have $X|D = \mathbf{q} \sim \mathcal{N}(\mathbf{m}, \Sigma)$, with

$$\mathbf{m} = \mathbf{m}_0 + \Sigma_0 A^T(\Gamma + A\Sigma_0 A^T)^{-1}(\mathbf{q} - \mathbf{m}_0), \tag{6}$$

$$\Sigma = \Sigma_0 - \Sigma_0 A^T(\Gamma + A\Sigma_0 A^T)^{-1}A\Sigma_0. \tag{7}$$

3 MCMC

Now that we have set up our problem, and we have built our posterior density, we can apply MCMC with the RWMA to obtain a sample from it. For doing this we will use the posterior distribution as defined in eq. 5 and we will compare the results with the values computed in eq. 6 and 7.

Notice that the posterior is proportional to the product of the likelihood with the prior. In this particular case we can compute the analytical form of the posterior with eq. 6 and 7, but this is not always the case. The acceptance probability in the RWMA algorithm is described in terms of a ratio (see section 3 in the first part about Markov chain Monte Carlo), hence using the posterior or the product of the likelihood with the prior is not relevant for the RWMA.

We set the proposal distribution as

$$\varepsilon \sim \mathcal{N}(0, S) \tag{8}$$

where $S \in \mathbb{R}^{n\times n}$ is a diagonal matrix with $S_{ii} = \sigma^2$, where σ^2 is the proposal variance.

4 Tasks

For this exercise you have to complete the following tasks:

1. Load the data contained in file *tc_Q_V.txt* and create the error array
2. Compute the mean and covariance matrix of the posterior as in eq. 6 and 7
3. Create three functions to compute prior, likelihood and their product
4. Starting from the code in the script *random_walk_metropolis.py*, set up the RWM algorithm.
5. Compare the results of RWM algorithm with those of the mean and variance of the posterior in eq. 6 and 7
6. Set $\sigma^2 = 0.06^2$, what is the length for the burn-in-period? Why?
7. What happens if we assume a noise level equal to 1? Comment it.

   What happens if we assume a noise level equal to 0.001? How do you have to set the proposal variance and the starting point for RWM? How is the burn-in-period affected? Justify your answer.
8. Set the noise level equal to 0.6 and vary the proposal variance (eg. $\sigma^2 = 0.01^2$ or $\sigma^2 = 3^2$). How does this affect the acceptance rate, the burn-in-period and in the end the result of RWM? Justify your answer.
9. Now let us assume another model:

   $$\mathbf{q} = a + b\mathbf{h} + c\mathbf{h}^2 + \eta \tag{9}$$

   with $\eta \sim N(0, \Gamma)$, with $\Gamma \in \mathbb{R}^{n\times n}$. How does the matrix A change? Compute all the tasks from 1 to 6 for the model in eq. 9. Which rules can you extrapolate for the choice of the proposal variance and of the length of the burn-in period?
10. Compare the models in eq 1 and 9. Which is the most appropriate in our case? Why?