P. Koumoutsakos, P. Hadjidoukas
ETH Zentrum, HG D1.1/G3
CH-8092 Zürich

Set 04 - Estimation of the posterior distribution

Issued: March 19, 2018
Hand in: March 26, 2018

Question 1: Finding optimal parameters analytically

Consider the problem of fitting a parameterized curve

$$y = \alpha_0 f^0(x) + \ldots + \alpha_M f^M(x)$$

to a set of data $D = (x_k, y_k)$, $k = 1, ..., N$. Here $f^i(x) : \mathbb{R} \to \mathbb{R}$ are user-selected known base functions and $\alpha_i \in \mathbb{R}$ are unknown parameters to be estimated using the data, $i = 0, \ldots, M$.

To account for model and measurement errors, assume the prediction error equation

$$y_k = \mathbf{f}_k^T \alpha + e_k,$$

where $\mathbf{f}_k = [f^0(x_k), \ldots, f^M(x_k)]^T$, $\alpha = [\alpha^0, \ldots, \alpha^M]^T$. The errors e_k are assumed to be i.i.d Gaussian: $e_k \sim \mathcal{N}(0, \sigma^2)$, where σ^2 is unknown. Assuming uniform priors, with large enough bounds, find:

a) The most probable values (MPVs) of α, σ^2

b) The Gaussian asymptotic approximation (GAA) of the posterior PDF

Question 2: Sampling the posterior distribution with rejection method

Assume that you are given 3 polynomial base functions $f^i(x) : \mathbb{R} \to \mathbb{R}$ (see Question 1): $f^0(x) = 1, f^1(x) = x, f^2(x) = x^2$. The data set $D = (x_k, y_k)$, $k = 1, ..., N$ is given in the file `data.txt`. We assume here that the error $\sigma^2 = 0.1$ is known.

We want to evaluate the posterior distributions of the parameters $y = \{\alpha_0, \alpha_1, \alpha_2\}$ computationally using the rejection sampling method. The algorithm to generate a single sample y is given by:

```
function REJECTION_SAMPLE(f, g, T)
    y ← RANDOM_G( )
    U ← RAND_UNIFORM(0, 1)
    ρ ← f(y)/Tg(y)
    if U < ρ then
        return y
    else
```

return REJECTION_SAMPLE(f, g, T)

where f is the pdf of the distribution we want to sample, g is the proposal distribution (easy to sample) and $T \in \mathbb{R}$ is such that $\forall x, Tg(x) \geq f(x)$.

a) Implement a serial version of the rejection sampling method in the provided skeleton code. For the proposal distribution g, use a uniform distribution with the bounds given in the code (see `lower_bound` and `upper_bound` variables). Visualize the posterior distributions for 10000 samples. What do you observe? How efficient is the rejection method, in terms of number of likelihood evaluations?

b) Parallelize your code using `OPENMP`. Experiment several methods in order to optimise the strong scaling of your code. Report the strong scaling of your implementation for up to 24 cores on Euler.

   Hint: on Euler, you may need to set the `OMP_PROC_BIND` environment variable to `TRUE`.