Lecturer: Dr. Dan Yamin
Teaching Assistant: Matan Yechezkel

# MCMC Assignment

| | |
|---|---|
| Form: | Code + comments |
| Language: | Preferable Python |
| Submission: | Code + Assignment report |
| Contact: | matanely@mail.tau.ac.il |
| Deadline for submission: | 18.06.2021 |

Goal: Understand MCMC theory and implementation methods

Each student will develop a mathematical model (e.g., linear regression, parabolic, exponential, or markovian, dynamic model or in any other form) to describe COVID-19 cases in Israel during a specific time frame and use MCMC to evaluate the posterior distribution of the model parameters.

**Submission:** Submission of the assignment will be made via Moodle by uploading the code file and the assignment report to the submission box. During the check, the code should run without errors.

**Part A – Modeling (30 pts):**

1. Download the COVID-19 daily new cases from קורונה - לוח בקרה (health.gov.il).
2. Develop a model that can capture the dynamics of COVID-19 daily new cases in the given period (longitudinal 60 days chosen randomly for each student). The model must include between 2-4 parameters. The model can be any reasonable model for this task (e.g., a dynamic model such as SIR model, regression, ML models, etc.). For example,

$$\mu_t = m_1 t + m_2 t \cdot \cos\left(\frac{2\pi(t+\varphi)}{\omega}\right)$$

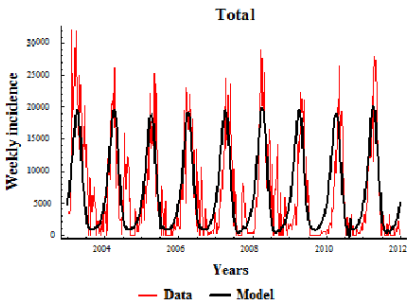
   where $\mu_t$ is the daily number of cases at time $t$ and $m_1$ and $m_2$ are the unknown model parameters.

3. Assume that the number of cases at time t, $Y_t$ is distributed according to a known distribution. For example,

$$Y_t \sim N(\mu_t, \sigma^2)$$

4. Calibrate model parameters with the data using **maximum likelihood estimation**. Write 1) the mathematical formula, 2) best-fit parameters.

5. Present on the same plot, the model fit, and the data for the given period. For example,



Further points for the report:

- The length of this section in the report should be up to 1 page at most! Points will be lost if exceeding one page.
- Explain the rationale for choosing this model for the given task.
- Provide a brief background for the model. Make sure to note model assumptions.
- Present the graph from section 5.
- Explain briefly the method used to calibrate the model and present the calibration results (parameters values, target function value).

**Part B – MCMC (50 pts):**

In this section, you are asked to implement all the required steps **from scratch**, unless explicitly said in these instructions (do not use any dedicated package for this task).

6. Write the code and run MCMC using the Metropolis-Hasting algorithm.
7. Evaluate the posterior distribution of model parameters.
8. Make sure the algorithm converged using the Gelman-Rubin test. Use 3 different randomized starting points $(\theta_0)$.

9. Present the following plot:
   9.1. Acceptance rate per iteration.
   9.2. Trace plot of 3 iterations for each parameter (each parameter in a separate plot).
   9.3. Posterior distribution of the parameters using one of the chains produced.

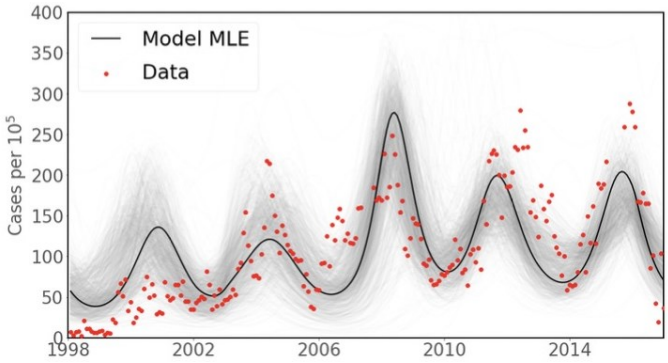
Further points for the report:

The length of this section in the report should be up to 2 pages at most.

- Present the initial algorithm parameters (scaling factor, covariance matrix, initial guess of model parameters).
- Present the graphs from section 9.
- For each graph (section 9), explain the results and their meaning in 1-2 sentences.
- Explain the results of the Gelman-Rubin test.

**Part C – 21 days projection (10 points):**

10. Project the upcoming 21 days using the MCMC. Specifically, sample 100 values of the parameters from the posterior distribution and run model projections.
11. Present projections in the same manner as shown below:




Namely, maximal likelihood estimate will be shown in bold black. Each of the 100 realizations will be presented in a light grey color. Data points will be shown in red.

Further points for the report:

The length of this section in the report should be up to 2 pages at most.

- Present the graph from section 11.
- Explain the results and the rational for your output.
- Why is it a good or bad representation of the reality?
- What could you do to improve your projection?

**Code documentation, report style (10 pts):**

Code documentation – code comments, proper naming, code structure.

Report style – Times New Roman or Calibri font, 12-pt font size, 1.5 line spacing.

Note –nowadays in both the industry and academic field – you **don't like to write, and we don't like to read!** Try to be as accurate and concise as possible.