Advanced Scientific Computing: Stochastic Methods for Data Analysis, Inference and Optimization

P. Koumoutsakos
323 Pierce Hall, 29 Oxford Street Cambridge,
MA 02138

Fall semester 2022

Set 2- Stochastic Simulations and Metropolis-Hastings Algorithm

Issued: September 28, 2022
Hand in: October 14, 2022 00:00

Question 1: Stochastic simulations of an epidemic model

The SEIR model describes the evolution of four populations (compartments) over time: susceptible individuals (S) can be infected in contact with infectious individuals (I) with rate β/N. Once infected, the susceptible individuals become exposed (E) and then become infectious (I) with rate μ. Furthermore, infected individuals become removed (R) at a rate γ. Assuming a well mixed population, the SEIR compartment model consists of the following reactions:

$$S + I \xrightarrow{\beta/N} E + I,$$
$$E \xrightarrow{\mu} I,$$
$$I \xrightarrow{\gamma} R.$$

a) Write a program that simulates a population of $N = 500$ individuals. Use $\beta = 0.5$ day^{-1}, $\mu = 0.05$ day^{-1} and $\gamma = 0.05$ day^{-1}. Start the simulation with only susceptible individuals except for 2.5% infected ones. Run your simulation until $t_{\text{end}} = 150$ days. Repeat the simulation 10 times and plot the 10 trajectories of the three compartments over time.

b) Write down the master equation for the above process. Use the master equation to derive the ordinary differential equation (ODE) for the expected compartment population sizes.

c) Solve the ODE numerically and compare the solution to the stochastic simulations. Vary the population size N. What do you observe when N is large?

d) For $N = 2000$, compare the ODE and 50 realizations of the SSA algorithm for an initial infectious population of 1. How do the two methods differ? Explain what you observe and why they differ.

e) Implement τ-leaping for the SEIR model and simulate the system described in part a) with different values of τ. Compare the total number of steps used by both SSA and τ-leaping methods. Note: make sure that the number of species does not become negative.

Question 2: Symmetric Random Walk

Consider a random walker on a lattice in d dimensions. The walker is initially placed at the origin. At each step, the walker randomly chooses one of the d directions and advances forward or backward (see fig. 1).

a) Write a program to simulate M walkers over N steps in d dimensions. The program should accept the following inputs:

- `N`, the number of steps to simulate per walker.
- `M`, the number of simulated walkers.
- `D`, the number of dimensions of the grid.

Note: Make sure that you store the trajectories of the simulated walkers, such that you can readily answer the following subquestions.

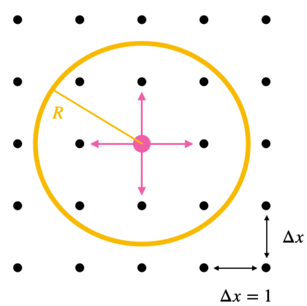

Figure 1: Sketch of the lattice in 2D. All walkers start at origin. At each step, a walker choses a random direction from a uniform probability distribution. Assume an infinite grid size.

b) Find the scaling law of the start-to-end Euclidean distance $d(n) = ||x_n - x_0||$ for different number of steps $n = \{0, .., N\}$ of the walk. For this, do a linear regression to find the parameters a and b of the model

$$\bar{d} = bn^a \,,$$

where $\bar{d}$ is the average start-to-end distance calculated from $M = 100$ trajectories of length $N = 1000$ steps. Plot the simulated start-to-end distance $d(n)$ against n. Include the fitted model $\tau^\star = bn^a$ to your plot and compare the solution for $d = 2$ and $d = 3$. Write down your observations in a few sentences.

c) Estimate the average number of steps $\bar{\tau}$ a walker needs to cross a given sphere with radius R around the origin. This is called the first passage time. Plot the average first passage time $\bar{\tau}$ and the standard deviation of τ against the radius $R = \{0, 1, 2, \ldots, 10\}$. Find the parameters a and b of the model

$$\bar{\tau} = bR^a \,.$$

Plot the fitted model $\tau^\star = bR^a$ for $R = \{0, 1, 2, \ldots, 10\}$ and compare the solution for $d = 2$ and $d = 3$. Write down your observations in a few sentences.

Note: You can reuse the calculations from subquestion b) and check the condition $d(n) \geq R$.

Question 3: Metropolis-Hastings Algorithm

Implement the Metropolis-Hastings algorithm to generate samples from a two-dimensional normal distribution with the mean at origin $\mu = (0, 0)^\intercal$ and the covariance matrix

$$\mathbf{\Sigma} = \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}.$$

Form a proposal distribution from two independent Cauchy distributions with median at the previous sample value and the scale parameter γ. Run the sampling for $\rho = 0$ and $\rho = 99/100$. For both runs plot 1000 samples and report the rejection rates. Why do more samples overlap in the second run?