Queen Mary
University of London

MTH739U / MTH739N / MTH739P: Topics in Scientific Computing 2017/18 – Coursework 2

Assignment date: Wednesday 29th November 2017

Submission Deadline: Friday 19th January 2018 at 23:55

You should attempt ALL questions. Marks awarded are shown next to the questions.

The coursework is due by **Friday 19th January 2018 at 23:55**. Please submit a report (preferably in pdf) containing the answers to the questions, complete with written explanations and printed figures. All figures should contain a title, axes labels, and a legend (if more than one curve are in the same figure). It is necessary to provide a sufficiently detailed explanation of all the original algorithms used in the solution of the questions (except for material explicitly discussed in the lectures, e.g. the Runge-Kutta method). You normally need to show that your program works using suitable examples. All the code produced to answer each question should be submitted in a single zip file, aside with the report. It might be useful to organise the code in different directories, one for each question. **Only material submitted through QMPlus will be accepted. Late submissions will not be considered. Plagiarism will not be tolerated. See the accompanying guidelines for additional information.**

Question 1 (10 marks). *Generating random numbers .*
Construct functions to obtain random numbers from the probability distribution detailed in the following, and test them using a sufficient number of samples, comparing the resulting histograms with the theoretical ones:

(a) Uniform distribution over the interval $[-2\pi, \pi]$. **[2]**

(b) Uniform distribution over the union of the three intervals $[1, 2] \cup [3, 4] \cup [5, 6]$. **[2]**

(c) Gaussian distribution with a given mean value μ and variance σ^2. Use either the Box-Muller method or the Marsaglia's polar method, and test your function by sampling from a Gaussian with $\mu = 12.5$ and $\sigma = 3$. **[2]**

(d) Continuous distribution with probability density function:

$$f(x; \lambda) = \lambda e^{-\lambda x}$$

for $x \geq 0$, where λ is a parameter provided by the user. Test your function by sampling from the three distributions obtained by setting λ respectively equal to 0.7, 1.5, and 3.5. **[2]**

(e) Continuous distribution with cumulative density function:

$$F(x) = \frac{1}{6}\left(x^2 + x\right)$$

for $x \in [0, 2]$. **[2]**

Question 2 (10 marks). *Monte Carlo integration.*
Implement five functions to compute Monte Carlo estimates for each of the following integrals, using $N = 5000$ samples uniformly drawn in the corresponding domain. Report in a table the values of your estimates and the corresponding variance.

(a)

$$I_1 = \int_{-2}^{2} \frac{1}{2} + \sin(x^3 - 4)\log(1 + |x|)dx$$

[2]

(b)

$$I_2 = \int_{0}^{10} \cos(x + 5)\frac{x^2 - 3x + 6}{x^2 + 3}dx$$

[2]

(c)

$$I_3 = \int_{2}^{8} \frac{\sqrt{\frac{1}{6}\left|\cos(x^2 - 5x + 6)\right|}}{\sin\left(\frac{x}{20} + 2\right)}dt$$

[2]

(d)

$$I_4 = \int_{-2}^{2} dx \int_{3}^{4} 3xy^2 \tan\left(\frac{x^2}{y}\right) dy$$

[2]

(e)

$$I_5 = \int_{0}^{1} dx \int_{0}^{1} \frac{\left|\sin\left(\sqrt{x^2 + 2y^2 + 5x + 1}\right)\right|}{\cos(xy)}dy$$

[2]

Question 3 (10 marks). *Importance sampling.*
Consider the integral:

$$I = \int_{0}^{300} 3x^{\frac{4}{3}}e^{-\frac{x}{4}}\mathrm{d}x$$

(a) Construct an Octave/Matlab function to obtain an estimate of I using the Mean-Value Monte Carlo method. Compute an estimate of I using $N = 10000$ points uniformly sampled in the integration interval, and report the corresponding standard error. **[3]**

(b) Determine a Monte Carlo estimate of I using Importance Sampling, using $N = 10000$ points sampled from the distribution with probability density function $p(x) = Ce^{-x}$, where C is a constant to be determined. Compute the corresponding standard error. **[3]**

(c) Compute and report in a table the values of the estimate and of the standard errors obtained by sampling $N = 10000$ points from the distribution with probability density function $p(x; \mu) = C_\mu \mu e^{-\mu x}$, where μ is respectively set equal to $\{0.1, 0.5, 1.5, 2.5\}$ and C_μ is a constant to be determined for each value of μ. **[4]**

Question 4 (12 marks). *An integral function.*
The function:

$$f(x) = -\int_0^x \log\left|2\sin\frac{t}{2}\right| \mathrm{d}t$$

is intimately connected with many special functions, including the polylogarithm function and the famous Riemann zeta function.

(a) Create an Octave/Matlab function which implements the Mean-Value Monte Carlo method, and use it to compute an approximation for $f(\pi/3)$ based on $N = 2000$ points uniformly drawn in $[0, \pi/3]$. Compute and report also the standard error of the estimate. **[4]**

(b) Use the Mean-Value Monte Carlo Method to approximate $f(x)$ in 100 equally-spaced points of the interval $[\pi, 2\pi]$. In particular, for each value of $x \in [\pi, 2\pi]$, compute a Monte Carlo estimate of $f(x)$ using $N = 500$ randomly drawn points. Use the results to produce an approximate plot of $f(x)$ for $x \in [\pi, 2\pi]$. **[3]**

(c) An alternative way of computing $f(x)$ is by using its Fourier series representation:

$$f(x) = \sum_{k=1}^{\infty} \frac{\sin kx}{k^2} = \sin x + \frac{\sin 2x}{2^2} + \frac{\sin 3x}{3^2} + \cdots \qquad (1)$$

Create an Octave/Matlab function that approximates $f(x)$ by truncating Eq. (1) to the first $n = 5$ terms, and produce another approximate plot of $f(x)$ using 100 equally spaced points in $[\pi, 2\pi]$. **[3]**

(d) Provide a quantitative argument to support the choice between the Monte Carlo method and the truncated Fourier series to compute $f(x)$, based on the comparison of the plots obtained in point (b) and (c) and on the relative efficiency and accuracy of the two methods. **[2]**

Question 5 (12 marks). *Random race.*
A random walk on the line is a generic discrete-time stochastic process whose trajectories start at time $t = 0$ at the position X_0 (the origin) and evolve according to the law:

$$X_{n+1} = X_n + \xi_n$$

where ξ_n is an increment/decrement sampled from a given probability distribution. We are interested in three specific flavours of random walks, as described below:

A) the discrete-space unbiased random walk, described by the probability distribution:

$$\begin{cases} P(\xi_n = +1) & = 0.5 \\ P(\xi_n = -1) & = 0.5 \end{cases}$$

B) the continuous-space random walk, where ξ_n is sampled (for each $n = 1, 2, \ldots$) from the probability density function:

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, \quad x \in \mathbb{R}$$

with $\mu = 0$ and $\sigma^2 = 2.25$

C) the continuous-space random walk, where ξ_n is sampled (for each $n = 1, 2, \ldots$) from the probability density function:

$$f(x) = \frac{\gamma}{\pi(x^2 + \gamma^2)}, \quad x \in \mathbb{R}$$

with $\gamma = 0.01$

In particular, we would like to determine which of the three random walks can move the farthest away from the starting point, and which one has the highest probability of remaining close to the starting point.

(a) Construct three distinct Matlab/Octave functions to simulate the three random walks A), B), and C) **[6]**

(b) For each of the three random walks in A), B), C), use the function you constructed to simulate $M = 500$ trajectories of length $N = 100$ for $X_0 = 0$. Then plot the histograms of the position X_{100} of the walker at the end of the trajectory, and of the maximum displacement of the walker from the starting point $X_0 = 0$. **[4]**

(c) According to the results of your simulations, which of the three random walks would you choose to ensure than the walker has a higher probability of remaining close to the starting point? Which of the three random walks can get the farthest away from the starting point? Motivate your answers. **[2]**

Question 6 (10 marks). *Random walk in 2 dimensions.*
Consider a uniform random walk on a plane, whose coordinates evolve according to the equations:

$$\begin{cases} X_n & = X_{n-1} + \xi_n \\ Y_n & = Y_{n-1} + \gamma_n \end{cases}$$

where ξ_n and γ_n are two *independent* random variables sampled from the same probability mass function:

$$\begin{cases} P(\xi = +1) & = P(\gamma = +1) & = 0.5 \\ P(\xi = -1) & = P(\gamma = -1) & = 0.5 \end{cases}$$

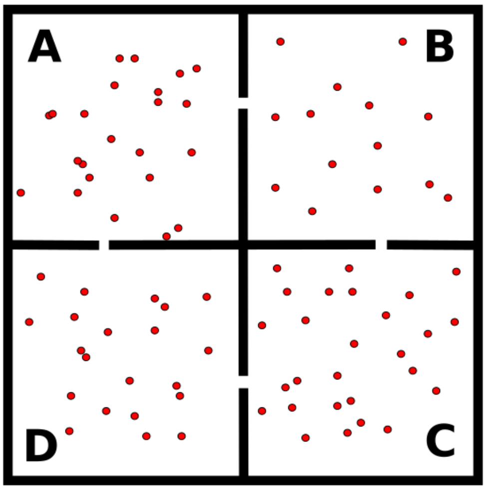
(a) Construct an Octave/Matlab function that takes as parameters a positive integer N and the initial values X_0 and Y_0, and simulates a trajectory of the uniform random walk on the plane starting at (X_0, Y_0) for N steps. **[3]**

(b) Use your function to simulate $K = 5$ distinct trajectories of the walk for $N = 200$, starting from $(X_0, Y_0) = (3, -1)$, and report them in a plot. **[2]**

(c) Simulate a trajectory of the random walk with $(X_0, Y_0) = (0, -2)$ and $N = 25000$. Count and report the number of time steps during which the position (X, Y) of the walker is such that $|X| > |Y|$. **[2]**

(d) Generate $K = 250$ trajectories of the walk of length $N = 170$, all starting from $(X_0, Y_0) = (0, 0)$. Plot the distribution of the X coordinate and the distirbution of the Y coordinate of the final position of the walker, and compare them with the theoretical ones. **[3]**

Question 7 (12 marks). *Diffusion of a gas.*
We are interested in studying the diffusion of a gas between the four adjacent boxes shown in the figure below:



Initially, box A contains $m_{A,0}$ molecules of the gas, box B contains $m_{B,0}$ molecules, box C contains $m_{C,0}$ molecules and box D contains $m_{D,0}$ of them, and the four boxes are isolated from each other (i.e., nothing can pass from one box to another). The total number of molecules in the four boxes is $m = m_t = m_{A,t} + m_{B,t} + m_{C,t} + m_{D,t}$, and remains constant in time. At time $t = 0$ the four boxes are put in communication as shown in the Figure (i.e. each box is connected to two adjacent boxes), and the gas contained in the boxes starts diffusing. This process can be considered as an extension of the classical Ehrenfest chain, where the system evolves according to the two following rules:

1) At each time step t, select one of the m molecules uniformly at random, and call it x;

2) take the molecule x from its current box, and move it to one of the two adjacent boxes, chosen with equal probability.

(a) Construct an Octave/Matlab function `four_boxes()` that takes two arguments, namely a vector m_0 containing the initial number of molecules in each of the four boxes (such that $m_0 = [m_{A,0}\ m_{B,0}\ m_{C,0}\ m_{D,0}]$) and a positive integer number N, and simulates the diffusion of gas in the four boxes for N time steps, starting from the initial condition m_0. **[4]**

(b) Use the function `four_boxes()` to generate a trajectory of length $N = 10000$ when the initial condition is $m_0 = [0, 400, 0, 0]$. Report in a single figure the time evolution of the four variables $m_{A,t}, m_{B,t}, m_{C,t}, m_{D,t}$ for $t \in [8000, 10000]$. **[3]**

(c) Simulate $M = 200$ trajectories of length $N = 5000$ with initial condition $m_0 = [150, 50, 150, 50]$. Plot the distribution of the number of molecules found in box B at time $t = 5000$ over the M trajectories. **[3]**

(d) Using the 200 trajectories obtained in point (c) above, provide an estimate of the expected value and standard deviation of the step t^* of a trajectory at which the number of molecules in box B becomes larger than 70. **[2]**

Question 8 (10 marks). *A network from the real world.*
The text file `network_500.txt` available at the following QMPlus URL:

https://qmplus.qmul.ac.uk/pluginfile.php/1251998/mod_resource/content/1/network_500.txt

contains the edgelist of an undirected network with $N = 500$ nodes, where nodes are labelled as $1, 2, 3, \dots 500$ and each edge is reported exactly once.

(a) Import the graph in Octave/Matlab, using a sparse matrix representation. **[2]**

(b) Compute and plot the degree distribution of the graph. **[2]**

(c) Compute the clustering coefficient of each node, and plot the corresponding distribution. **[2]**

(d) Plot in a figure the value of the clustering coefficient of each node versus the corresponding value of node degree. **[2]**

(e) Compute the value of the Spearman's rank correlation coefficient between the degree and the clustering coefficient of the nodes of the graph. **[2]**

Question 9 (14 marks). *On random graphs, and how they get connected.*
An Erdös-Renyi random graph (model A) on N nodes is obtained by sampling uniformly at random K of the $\binom{N}{2}$ possible pairs of nodes and connecting each sampled pair with an edge, avoiding the creation of multiple links.

(a) Implement an Octave/Matlab function to sample random graphs from the Erdös-Renyi model A with N nodes and K edges. **[3]**

(b) Use the function implemented in point (a) to construct an Erdös-Renyi random graph with $N = 10000$ nodes and $K = 6000$ edges. Compute and plot the degree distribution of the obtained graph, and compare it with the theoretical one. **[3]**

(c) Use the function implemented in point (a) to construct a series of Erdös-Renyi graphs with $N = 1000$ nodes and 80 increasing values of K in the range $[50, 4000]$. Plot the value of the largest node degree observed in each of those graphs as a function of the average degree of the graph. **[4]**

(d) For each of the 80 graphs obtained in point (c), compute the relative size of the largest connected component S (i.e., the number of nodes contained in the largest connected component divided by the total number of nodes in the graph N) and the average size $\langle s \rangle$ of all the connected components of the graph *except* the largest one. Plot the values of S and $\langle s \rangle$ as a function of the average degree of the graph. **[4]**

End of Paper.