Modeling and simulating malaria

Author(s): Seyla Wachlin, Rick Quax

Note

Please **work in pairs** for this assignment.

Problem statement

Malaria is a severe disease that is caused by a parasite. The parasite is spread from one infected human to the next by a mosquito. Mosquitoes are born uninfected, but in the course of their lifetime they need to obtain several blood meals. Once they bite an infected person, they become infected themselves and can spread the disease next time they bite. See e.g. http://en.wikipedia.org/wiki/Malaria, http://www.who.int/mediacentre/factsheets/fs094/en/ and http://en.wikipedia.org/wiki/Anopheles. For this assignment you will be creating a model for the spread of the disease in a human population and you will study its behaviour, and a way to prevent or mitigate it.

Implementation of the model: a reference

The following is a reference model which you may choose to implement and study. You are also free to choose your own design of a model for Malaria infections, as long as it makes sense (and you can explain it).

In this simulation you will deviate a little from the exact definition of cellular automata, and also add the notion of probabilities. We will consider a two-dimensional grid. Populate the grid with a number of humans, at most one per cell. Each human can be in one of the following states: susceptible, infected, immune, or dead (and removed). Assume that the total human population is stable, so we will assume that a new human is added somewhere ('a baby matures') every time someone dies. People can die of malaria, or from other causes.

Also create a population of mosquitoes. As opposed to humans, mosquitoes can move over the grid freely and be co-located (multiple mosquitoes in one cell), performing a random walk, from one cell to the next. Mosquitoes can be hungry, or not (they only bite when hungry, and in the same cell as a human), and infected, or uninfected.

Lastly, you should implement one or more hypothetical prevention methods of your choice, such as the use of mosquito nets or the use of drugs to prevent people from catching malaria, or any other thing that you may think of. For more examples see https://en.wikipedia.org/wiki/Malaria#Prevention.

Fitting the model parameters to real data

Your model will have many parameters, such as the infection probability (both mosquito-human and human-mosquito), death rates, the initial distributions to start each simulation, etc. It would be wise to organise all your parameters in one data structure (such as a Python dictionary) to keep track of them. Your first order of business is to find a suitable value for each of these parameters.

What does it mean for a particular choice of parameter values to be 'suitable'? It means that your model reproduces facts that we measure in real life. The most straightforward fact to fit to is the so-called 'prevalence' of malaria among humans: the percentage of humans that are infected with malaria. See e.g. `http://www.internewskenya.org/dataportal/categorylist/11`. This varies from region to region, so simply choose a location that you want to model (e.g., in Kenia, close to a lake?). Eventually your assumptions should result in one number to fit, e.g., a stable prevalence of 20% (or whatever other number you find[1]). By 'stable' prevalence we mean that from the start of the simulation the prevalence may change a lot, but after some time it should settle to around a certain value[2], which for 'suitable' parameter values should be around the real value of e.g. 20%.

Optionally, you can fit additional real facts as well to obtain even more realistic parameter values. One idea would be the so-called 'vector density' of malaria-infected mosquito's, but please feel free to choose anything else.

How to fit

One way to fit the parameter values is to fit them by hand. That is, you manually change a parameter value, run the simulation for some large number of time steps, see if the stable prevalence has gotten closer to the real target value or not, then change again some parameter, and so on, until finally you are satisfied with the result.

A more advanced (and optional) way would be to automate this fitting process. Parameter optimisation is a research field on its own and we will not go into detail here. But one simple way is using random sampling. First, for each parameter you define the set (or range) of allowed values. Then you randomly generate a set of parameter values, run the simulation, and automatically estimate the 'error' that the simulation has, such as the absolute difference between the stable prevalence in the simulation and the real value.

As a result, it would be insightful to show a single plot that demonstrates how a representative simulation of your model evolves over time relative to the real fact, e.g., how its prevalence settles to the 20% mark.

[1] Just be sure to cite the source where you get your information from.

[2] If your prevalence keeps changing a lot, consider increasing the grid size and the numbers of humans and mosquitoes to get better statistics.

Experiment(s) and analysis

Having implemented your model and fitted its parameters to make it realistic, it is now time to analyse your model to gain further insights into the spread of malaria. Analyse the effects of your implemented prevention method(s). For example, study the effects of using pesticides on $x\%$ of the mosquito population or study the effects of giving mosquito nets to $y\%$ of the human population, or anything else that you may think of. (Can you eradicate the disease altogether?)

You are free to perform any study on this model. Just make sure that you make one to three insightful plots that summarise the main findings of your experiments. Explain in text what you studied, why it is interesting, and what you found.

Assignment

Please submit to Canvas:

1. A short written report with the following main headers:
   (a) *Model definition and implementation.*
   (b) *Fitting the model parameters.* Suggestion: one figure showing how a single run of the simulation proceeds over time in terms of fractions of infected/healthy/dead over time, and how it settles to the real (target) value eventually. Also: a table with all parameter names and their values.
   (c) *Experiments and analysis.* one to three insightful figures.
2. Your code, functional and commented. Your code is required to run successfully. When run, it should output all figures used in your document, either on screen or in the same working directory (named e.g. as `fig1.png`).

It is important to note that your experiments and analyses should be summarised in one to three insightful figures. Each figure should have a caption which describes what it shows and how it was computed, such that it is more or less stand-alone and can be understood by itself. Your report's purpose is to logically connect the figures and motivate them, and draw a main conclusion from them. Each plot may be a combination of subplots which are logically connected. Please keep your report concise, meaning to-the-point, and preferably limited to about four pages.

Grading

- Report document: 70%

  Does the report have a clear structure and a logical flow; are the figures insightful and well-described; are the effects of the prevention method(s) clearly set out; are main conclusions drawn; etc.

- Python code: 30%

Does the code work; is it well-structured and commented; does it output the same plots as in the report document; which prevention method(s) are implemented; etc.

Your code will be tested for plagiarism using special-purpose software. Also, in case your report contains figures which your code cannot believably produce we will investigate the possibility of plagiarism.

This assignment will count 50% towards your grade for the CA practical assignments (15% towards final grade).

Deadline: Friday, 8 March 2019, 23:59.