Modelling Infectious Disease Spread

Jacques Chester 20304893

This report is submitted as partial fulfilment
of the requirements for the Honours Programme of the
School of Computer Science and Software Engineering,
The University of Western Australia,
2010

CHAPTER 1

Introduction

Infectious disease has been a powerful agent of suffering and change throughout history and continues to be so today. The fate of many nations has turned on the spread disease from person to person. In the ancient world epidemics were frequently seen as instruments of divine wrath or justice. The germ theory of disease, established and popularised in the modern Western period by Louis Pasteur, demonstrated that providence was not at work in plagues; merely reproducing microorganisms. This did not, of course, reduce their deadliness. But the introduction of simple hygiene, isolation of infected patients, and most of all immunisation, have prevented millions of unpleasant deaths or afflictions since Pasteur's work in the 1860s. The World Health Organisation estimates that in 2002 alone, two million lives were saved by immunisation [7].

Immunisation serves a double purpose: it prevents the immunised person from becoming infected, and it prevents them from acting as an agent of further infection. In this way, immunisation can prevent epidemics and even (in the case of Smallpox) eradicate diseases altogether. With mass immunisation under attack from various pseudo-scientific quarters, it is more important than ever to demonstrate its efficacy in preventing disease.

1.1 Modelling Epidemics

Models of infections disease spread are often used to serve public health ends.

Firstly, models can explore situations that experiments cannot. It is unethical to perform human trials of epidemics by infecting patients and placing them in the general population. Disease-spread models, in conjunction with animal trials, allow investigations of how different parameters of disease and population interact to cause certain disease spread patterns. The advent of computers has allowed ever-larger models to be constructed and investigated to improve understanding of how to identify, prevent and manage epidemics.

Secondly, these tools can be applied in "real time". When a new disease emerges, models can be used to predict its future behaviour based on revealed parameters. The results of the model can then calibrate the public health response to maximise population health, minimise harm and potentially prevent unnecessary outlay from the public purse.

1.1.1 Mean-field models

Mean-field models of disease spread express behaviour in terms of systems of equations. The Kermack-McKendrick model[4], the best-known of these, proposes three equations which relate time to Susceptible (S), Infected (I) and Recovered (R) patients. For this reason these models are often called "SIR models".

SIR models have the advantage that they can quickly provide a 'high-level' view of how an epidemic might behave over time. However, they can exclude factors that might or might not affect the real-world result. Since Anderson and May[1] reintroduced the Kermack-McKendrick SIR model, there have been many elaborations based on adding or subtracting various parameters[2].

1.1.2 Agent-based models

Agent-based models work by using finite automata to simulate members of a population. Rather than having an equation dealing with *e.g.* Susceptibility on a population-wide basis, each agent is a finite state machine that is discretely in one state only.

An advantage of agent-based models is that they lend themselves neatly to visualisations of disease spreads. They can also be scaled up by having agents distributed to multiple cores or even multiple machines in a network. Agent-based models also lend themselves to statistical sampling of parametric space through the Monte Carlo method.

CHAPTER 2

Responses to Questions

2.1 Modelling Code

The modelling code submitted for this assignment is written in Ruby. Individual cells are modelled as state machines using the Acts As State Machine library[3]. Cells belong to a neighbourhood (in this case a 9-cell Moore neighbourhood), which they query to detect illness in their vicinity. Cells also belong to a lattice, providing a two-dimensional grid in which diseases can spread. Lattices are finite and do not form a torus. Neighbourhoods are truncated as required to fit within the boundaries of a lattice.

Mathematically it is simple to distinguish between a state S_t and S_{t+1} .In programming terms this is slightly more difficult: the system behaves differently if all cells can update their own state asynchronously. Therefore some mechanism is required to perform simulation steps for all cells synchronously.

To recreate the mathematical behaviour of states with successor states, two lattices are used. Any given cell's neighbourhood is actually composed of cells from the *other* lattice. Each lattice is called to update its simulation in turn. Thus each of the two lattices alternate between being the current state and the successor state, a technique called "tick-tock simulation" (Figure 2.1).

During simulation, the code records results to an SQLite database[1]. Multiple simulations can be performed using the Monte Carlo method (see the command line help for more details).

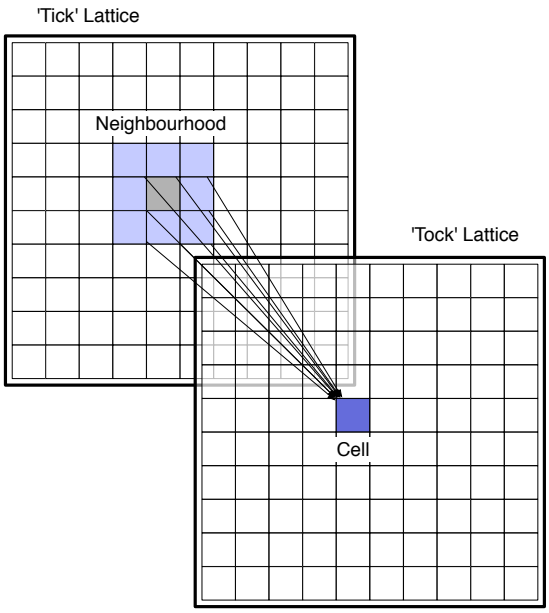

Figure 2.1: An illustration of the relationship between cells, neighbourhoods and dual lattices

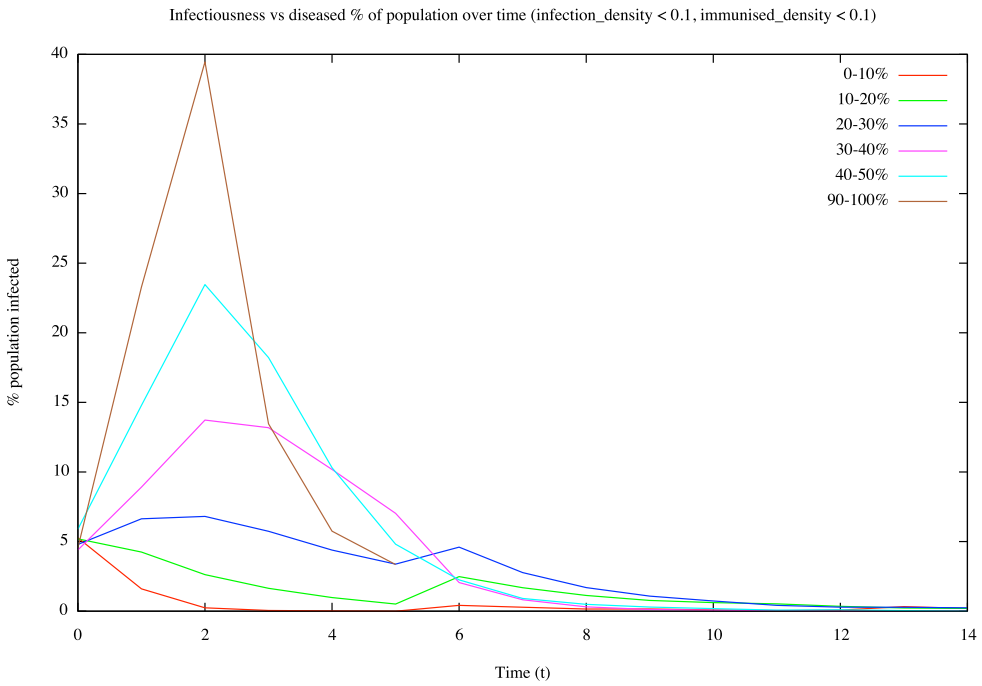

Figure 2.2: % of population infected over time at different infectiousness rates

2.2 Setting P1=0.5

Epidemics in SIR models follow a left-skewed distribution. Infection rises rapidly and then falls off as infected agents run out of susceptible agents in their neighbourhoods. Figure 2.2 shows how several range values of P1 (probability of infection) change the behaviour of the epidemic.

In particular, infection probabilities below the 20-30% range do not cause epidemics. Infections remain localised and the majority of the population does not come into contact with the disease. The infection "slow burns" and eventually disappears.

In contrast, infection probabilities above 30% begin to display the left-skewed curve predicted by an SIR model. Infection rises quickly, falls quickly and then tapers off. This effect is most visible in the 90-100% range, where infection spikes up to T_2 and then falls away almost as rapidly.

2.3 Immunising 50% of the population

Trials with initial immunisation rates between 49% and 51% show a strong dampening effect on the spread of disease (Figure 2.3, set to the same scale as Figure **??**). Even when the disease is extremely infectious, the disease spreads relatively slowly and only infects a smaller portion of the population. With immunisation rates below 10%, epidemic behaviour emerges above 30%. With prior immunisation of 50%, epidemic behaviour does not emerge until the 50-60% infection probability range.

2.4 P1's effect on P_{th}

P1 - the probability of death after contracting a disease - does not seem to have a large effect in the model (Figure 2.4). Infection rates are tightly clustered along a SIR-like left-skewed distribution, regardless of whether cells live or die.

This may or may not reflect real-world epidemics. Very lethal diseases can pass through small communities quickly, thus preventing spread to larger communities by killing or immunising all potential hosts in the smaller group first. This may be a property of uniform neighbourhood model; as different behaviour

[1]There is a known bug in the code to do with the storage of data from multiple invocations of the script.

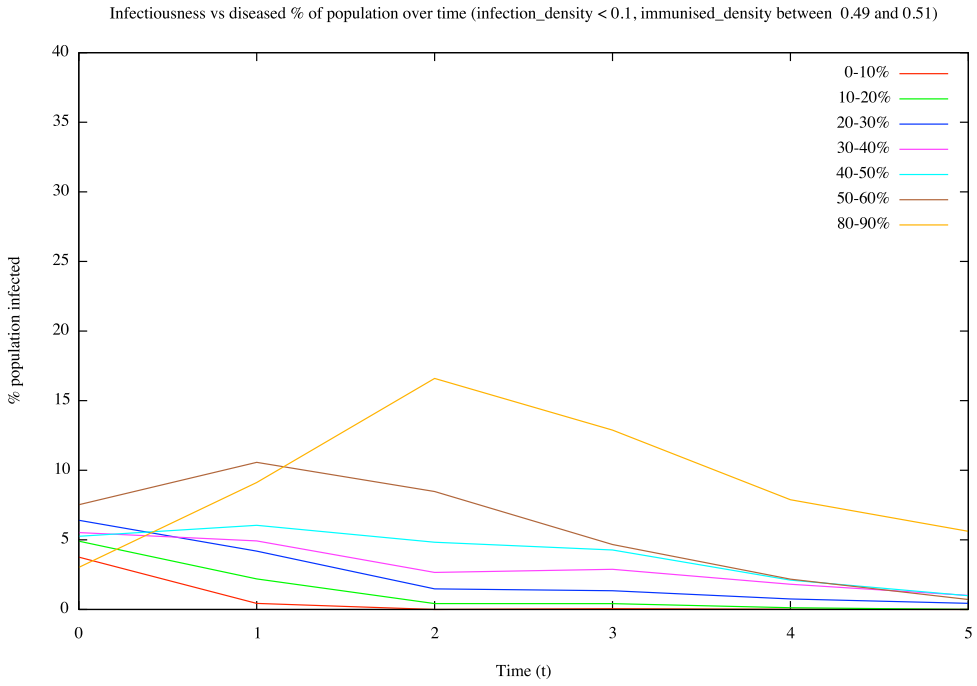

Figure 2.3: % of a 50% immunised population infected over time at different infectiousness rates

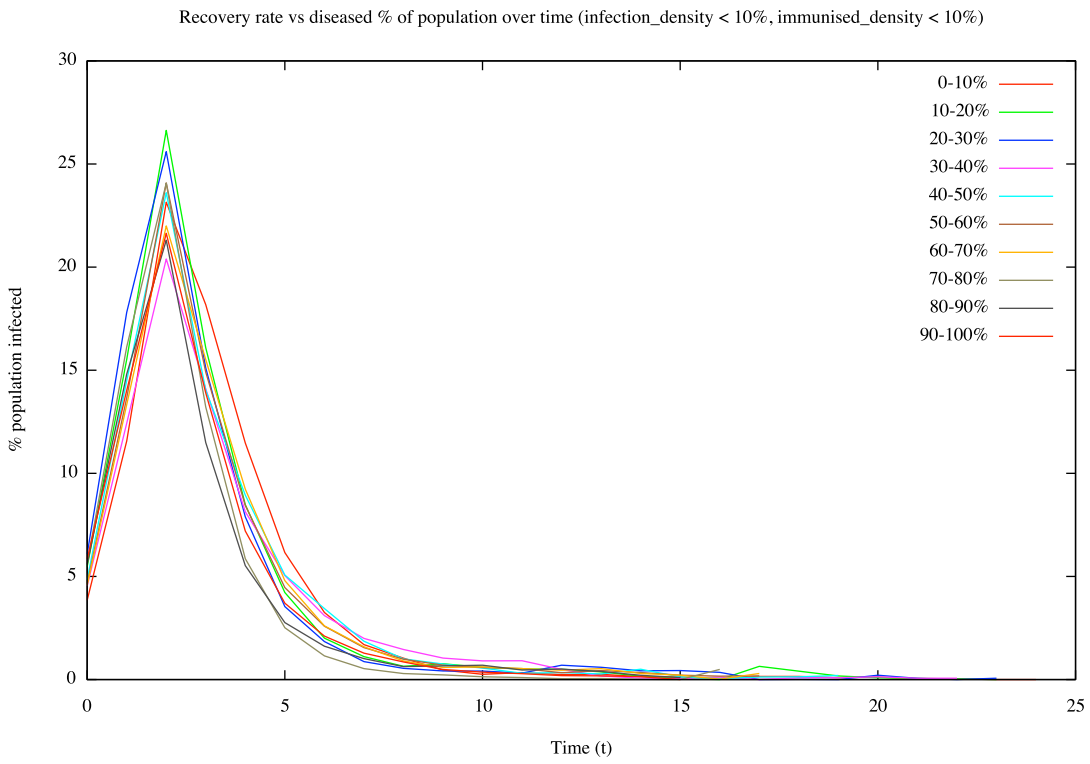

Figure 2.4: % of population infected over time at different recovery rates

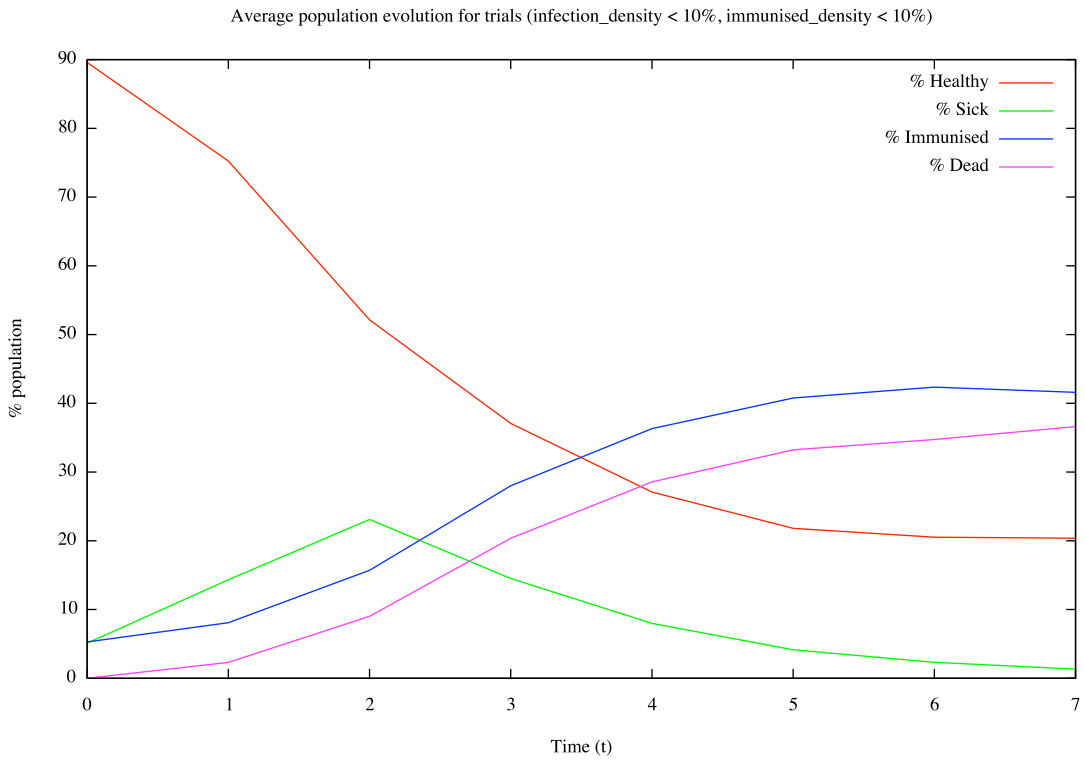

Figure 2.5: Average population evolution for multiple trials

could be expected in isolated subgraphs of a population vs the population as a whole (see the discussion on modifications).

2.5 Population plots

Out of 22,839 Monte Carlo trials performed for this assignment, 219 trials have initial infection rates below 10% and immunisation rates below 10%. For these the model behaves on average in line with a standard SIR model (Figure 2.5).

The use of Monte Carlo trialling allows for ad hoc investigation of different relationships - in this case through SQL queries - as shown in previous sections on immunisation, recovery probabilities and infection probabilities. It may also lend greater statistical confidence that the results are more robust over a range of pseudo-random scenarios.

2.6 Model modifications

One flaw with a lattice-based agent model is that neighbourhood size is identical for all agents. This does not correspond with social contact between humans. For example, *The Economist* reports research on Facebook's social graph by Cameron Marlow[5]. Marlow finds that the average Facebook user has 120 'friends', with some friend graphs rising as high 500 members. Interactions occur at different levels of activity, meaning that time in contact will vary. So too, presumably, would infection probability.

This means that an agent model should allow for variably-sized neighbourhoods, perhaps following a power-law distribution. It should further allow for each cell to have different infection probability. Simulation would change from being lattice-based to network- or graph-based. Lattices would be replaced with sets of cells related through the variably-sized neighbourhoods. The tick-tock simulation technique would be retained.

On the downside, this complicates graphical display of simulations, though this is a solvable problem (see the STEM project for an example.[6]).

CHAPTER 3

Conclusion

Agent-based models provide a simple-to-understand mechanism for examining the behaviour of disease spread in a population under a variety of conditions. They also lend themselves to relatively straight-forward Monte Carlo simulations, evenly sampling the space of all potential parameter values, to extract average outcomes under various conditions. The software developed for this assignment provides a finite-state machine model with configurable neighbourhoods and automatic Monte Carlo simulation support.

The agent-based model performs comparably to the SIR model, affirming that agent-based models may sometimes be used in lieu of mean-field models and *vice versa*.

Bibliography

[1] Anderson, R., and May, R. Population biology of infectious diseases: Part i. *Nature 280* (1979), 361–367.

[2] Anonymous. Compartmental models in epidemiology. *Wikipedia.org*.

[3] Barron, S. Acts As State Machine.

[4] Kermack, W., and McKendrick, A. A contribution to the mathematical theory of epidemics. *Proceedings of the Royal Society A 115* (1927), 700–721.

[5] Kluth, A. The size of social networks. *The Economist* (2009).

[6] The Eclipse Foundation. The Spatiotemporal Epidemiological Modeller (STEM) Project.

[7] World Health Organisation. Fact sheet 288: Immunization against diseases of public health importance, 2005.